CLASS II DNA METHYLTRANSFERASES OF ZEA MAYS

FIELD OF THE INVENTION

The present invention relates to nucleic acid and amino acid sequences which encode class II DNA methyltransferases. The present invention further relates to methods of using the nucleic acid and amino acid sequences described herein to stabilize transgene expression in transgenic plants, to alter the yield or biochemical qualities of plants and to silence targeted genes in plants *in vivo*.

BACKGROUND OF THE INVENTION

The information content of a primary DNA sequence can be enhanced by the addition of a methyl group to the ring structure of cytosine or adenine residues (Finnegan, E.J., et al., *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49:223-47 (1998)). The chemical modification of DNA is known to affect protein-DNA interactions. Specifically, in prokaryotes, methylation of DNA prevents cleavage by the cognate restriction endonucleases. *Id.* In higher eukaryotes, cytosine methylation can inhibit binding of regulatory proteins and methylation of promoter and coding sequences of genes can repress transcription, both *in vitro* and *in vivo*. *Id.* Methylation of DNA has been implicated in the timing of DNA replication, in determination of chromatin structure, in increasing mutation frequency, as a causal agent for some human diseases, and as a basis for epigenetic phenomena. *Id.*

Eukaryotic genomes are not methylated uniformly, but instead contain specific methylated regions, with other domains remaining unmethylated (Martienssen, R.A., et al., *Current Opinion in Genetics and Development*, 5:234-242 (1995)). The enzymes that transfer methyl groups to the cytosine ring are cytosine-5-methyltransferases (hereinafter referred to as "DNA methyltransferases") and have been characterized from a number of eukaroytes. All characterized eukaryotic DNA methyltransferases exhibit little primary sequence specificity *in vitro* other than the short canonical symmetrical sites methylated which are CpG in animals, and CpG and CpNpG in plants (where N stands for any nucleotide). Mammalian and plant

genomes contain methylation-free GC-rich zones, or CpG islands, which are frequently associated with the 5' regions of housekeeping genes. *Id.*

In plants, DNA methylation is necessary for normal development. For example, Arabidopsis having reduced levels of DNA methylation demonstrate a range of abnormalities, including loss of apical dominance, reduced stature, altered leaf size and shape, reduced root length, homeotic transformation of floral organs and reduced fertility (Finnegan, E.J., et al., *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49:223-47 (1998)). Moreover, Arabidopsis plants in which methylation had been reduced by at least 70% became infertile after four to five generations of selfing. *Id.* A comparable reduction in DNA methylation is embryo lethal in mammals. *Id.*

Two classes of DNA methyltransferase enzymes have been cloned in plants (Finnegan, E.J., et al., *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49:223-47 (1998)) - class I and class II. Class I enzymes include MetI and MetII from Arabidopsis (Finnegan et al. *Nucleic Acids Res.*, 21(10):2383-2388 (1993); Nebendahl, et al., *Gene* 157(1-2):269-272 (1995)), Met1-5 and Met2-21 from carrot (Bernacchia, G et al., *Plant Physiol.* 116:446-446 (1998)), C-5 MTase from tomato (Bernacchia, G et al. *Plant J.*, 13(3):317-330 (1998)), and C-5 MTase from pea (Pradhan et al., *Nucleic Acids Res.*, 26(5):1214-1222 (1998)). Class II sequences have been detected in many species with a defining characteristic of the presence of an embedded chromodomain (Rose et al., *Nucleic Acids Res.*, 26(7):1628-1635 (1998)). The only full-length class II sequence is Cmt1 from Arabidopsis (Genbank #AF039364).

Class I enzymes are homologous to dnmt1 from mice (Bestor, T., et al., *EMBO J.*, 11(7):2611-2617 (1988)), the first cloned DNA methyltransferase. A knockout of dnmt1 in mice resulted in lethality during embryogenesis (Li et al., *Cell*, 69(6):915-926 (1992)). Dnmt1 has been used as a model for all class I enzymes though it has not been proven whether this is appropriate in plant systems. Antisense expression of MetI in Arabidopsis resulted in numerous developmental abnormalities (Finnegan et al., *Proc. Natl. Acad. Sci. U.S.A.*, 93(16):8449-8454 (1996)). Class I enzymes are thought to function as maintenance enzymes, though proteolytic cleavage could create de novo enzymes (Bestor, T.H., *EMBO J.*, 11(7):2611-2617

(1992)). CpG activity has been shown for dnmt1 in mice and humans. In peas it was found that pea C-5 MTase expressed in baculovirus displayed both CpG and CpNpG activity (Pradhan et al., *Nucleic Acids Res.*, 26(5):1214-1222 (1998)). In general, class I enzymes have a high level of expression in tissues that are actively dividing and are expressed at lower levels or silent in mature tissues.

There is little known regarding the function of class II enzymes. Cmt1 was detected as an Arabidopsis genomic sequence based on sequence homology to other methyltransferases. The C-terminal region contains the conserved methyltransferase domains and a chromodomain. The N-terminal region is much shorter than the N-terminal region of class I enzymes. Several commonly used ecotypes of Arabidopsis contain an allele of Cmt1 which is interrupted by a transposon insertion. These Cmt1 knockouts do not have any detectable phenotype. No other research has been published on the function of class II enzymes. Cmt1 is expressed only in floral tissues at very low levels. Degenerate PCR has been used to show the presence of Cmt1 homologs in a number of other plant species (Rose et al., *Nucleic Acids Res.*, 26(7):1628-1635 (1998)). In addition to finding homologs in other species, two sequences with similarity to Cmt1, Cmt2 and Cmt3, were identified in the Arabidopsis.

DNA methylation provides a mechanism for the mitotic propagation of epigenetic states. Epigenetic lineage-dependent patterns of gene expression have been studied the most in the germline and in somatic cell lineages in multicellular eukaryotes (Martienssen, R.A., et al., *Curr. Opin. Genet. and Develop.*, 5:234-242 (1995)). For example, in mice, the parentally imprinted genes *H19* and *Igf2r* are expressed in the embryo only when they are inherited via the female gamete. *Id.* In contrast, the *Igf2* gene is expressed only when inherited via the male gamete. *Id.* The human homologs of the *Igf2* and *H19* genes are linked and parentally imprinted as in the mouse. *Id.* Parental uniparental disomy for this chromosomal region (11p15) is associated with Beckwith-Wiedemann syndrome, which is believed to result from overexpression of *Igf2*. *Id.* In addition to overgrowth of certain organs, Beckwith-Wiedemann syndrome patients have a 700-fold predisposition to Wilms' tumor, and loss of heterozygosity in this region is found in many other tumors as well. *Id.* It has

also been shown that 60-70% of Wilms' tumor patients have biallelic expression of *Igf2, H19*, or both in tumor tissue, resulting from loss of imprinting rather than loss of heterozygosity. *Id.*

In plants, epigenetic changes in gene expression are considered to be easier to observe than in animals since there is little cell migration and clonal lineages stay together. *Id.* Moreover, because in plants the germline arises relatively late in development, many somatically variegated phenotypes can be followed into the next generation and are heritable to greater or lesser extents. *Id.* Parental imprinting of gene expression was first observed in plants at the *R* locus in maize. *Id.* Certain alleles condition a mottled phenotype in the alerone layer of the extra-embryonic endosperm when inherited paternally, but cause a fully colored phenotype when inherited maternally. *Id.* Genetic studies of modifier loci have revealed that it is the maternally inherited *R* allele that is imprinted to a high level of expression. *Id.* High levels of R expression correlate with demethylation of sites in the transcribed region in the maternally inherited allele. *Id.*

Plants transformed with additional copies of endogenous genes or with multiple copies of a foreign or exogenous gene (these endogenous and exogenous genes are often referred to as "transgenes") frequently display epigenetic inactivation. This phenomenon is known as "gene silencing" or "co-suppression". There are two types of "gene silencing" or "co-suppression". The first is "transcriptional silencing". In "transcriptional silencing", RNA production from the introduced transgene is repressed. The second type of "gene silencing" is "posttranscriptional silencing". In "posttranscriptional silencing", transcripts do not accumulate in the cytoplasm even though transcription rates are comparable with or are higher than those in cells where transcripts do accumulate.

Transcriptional silencing is associated with transgene methylation, particularly in the promoter (Finnegan, E.J., et al., *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49:223-47 (1998)). Posttranscriptional silencing, which affects both transgenes and homologous endogeneous genes, is also associated with transgene methylation, but within the coding sequence rather than the promoter. *Id.* It is believed that both

forms of gene silencing reflect normal, cellular defenses against invading or mobile DNAs. *Id.*

Currently, two classes of methyltransferase genes have been cloned in maize. The class I clone homolog is referred to as Zmet1 and the class II homolog Zmet2. The Zmet1 is a class I enzyme that was cloned by Paula Olhoft and Ron Phillips at the University of Minnesota. FIG. 4 is a summary of the major classes of 5-cytosine methyltransferases from mammals, *Arabidposis* and maize. The present invention herein relates to zmet2a and zmet2b methyltransferases.

SUMMARY OF THE INVENTION

In one embodiment, the present invention relates to an isolated and purified *Zea mays* zmet2a methyltransferase nucleic acid sequence. Specifically, the isolated and purified *Zea mays* zmet2a methyltransferase nucleic acid sequence of the present invention hybridizes to the nucleic acid sequences shown in FIG. 1A and 1B under stringent conditions. The zmet2a methyltransferase nucleic acid sequence encodes the enzyme zmet2a methyltransferase. The amino acid sequences for zmet2a methyltransferase is shown in FIG. 2A and FIG. 2B.

In another embodiment, the present invention further relates to recombinant expression cassettes comprising the isolated and purified zmet2a nucleic acid sequence described herein. Preferably, the recombinant expression cassettes further contain a promoter sequence and a polyadenylation signal sequence. The promoter sequence can be operably linked to the zmet2a nucleic acid sequence. The zmet2a nucleic acid sequence is operably linked to the polyadenylation signal sequence. Any promoter sequence can be used in the recombinant expression cassette, such as, but not limited to a constitutive or tissue specific promoter.

In another embodiment, the present invention also relates to a recombinant expression cassettes comprising one or more heterologous nucleic acid sequences. Such recombinant expression cassettes further contain a promoter sequence from the zmet2a nucleic acid sequence and a polyadenylation signal sequence. The promoter sequence is operably linked to the heterologous nucleic acid sequence. The

heterologous nucleic acid sequence is operably linked to the polyadenylation signal sequence. Any heterologous promoter sequence can be used in this recombinant expression cassette.

In a further embodiment, the present invention also relates to bacterial cells comprising at least one of the recombinant expression cassettes described herein. The bacterial cells can be *Agrobacterium tumefaciens* or *Agrobacterium rhizogenes*.

In a further embodiment, the present invention further relates to transgenic plant cells and transgenic plants containing the recombinant expression cassettes described herein. Monocotyledonous or dicotyledonous plant cells and plants can be transformed with the hereinbefore described recombinant expression cassettes. Plants which can be transformed with the recombinant expression cassettes of the present invention include, but are not limited to, *Zea mays*, *Oryza sativa*, *Secale cereale*, *Triticum aestivum*, *Daucus carota*, *Brassica oleracea*, *Cucumis melo*, *Cucumis sativus*, *Latuca sativa*, *Solanum tubersoum*, *Lycopersicon esculentum*, *Phaseolus vulgaris*, *Brassica napus*, etc. The present invention also relates to seed resulting from the transgenic plants of the present invention.

In a further embodiment, the present invention further provides methods of reducing or altering methyltransferase activity in a transgenic plant in order to increase transgene expression stability and/or to improve the yield or biochemical qualities of a plant as well as a method of silencing targeted genes in a plant *in vivo*. These methods comprise introducing into a plant a recombinant expression cassette comprising an appropriate plant promoter operably linked to a zmet2a methyltransferase nucleic acid sequence described herein in either the sense or antisense direction.

In a further embodiment, the present invention relates to an isolated and purified *Zea mays* zmet2b methyltransferase nucleic acid sequence. The zmet2b methyltransferase nucleic acid sequence of the present invention can be isolated using an isolated and purified partial *Zea mays* zmet2b methyltransferase nucleic acid sequence. The isolated and purified partial *Zea mays* zmeb2b methyltransferase

nucleic acid sequence can be used as a probe to isolate the zmet2b methyltransferase nucleic acid encoding zmet2b methyltransferase. Preferably, the isolated and purified partial *Zea mays* zmet2b methyltransferase nucleic acid described herein hybridizes to FIG. 23 under stringent conditions. The partial zmet2b methyltransferase nucleic acid sequence described herein encodes a portion of zmet2b methyltransferase. The partial amino acid sequence of zmet2b methyltransferase is shown in FIG. 24. The zmet2b methyltransferase nucleic acid sequence can be used in recombinant expression cassettes in the same manner as the isolated and purified zmet2a nucleic acid sequence described herein. Such recombinant expression cassettes can be used to create transgenic plants containing these recombinant expression cassettes. Additionally, the zmet2b methyltransferase nucleic acid sequence can be used to reduce or alter methyltransferase activity in transgenic plants in the same manner as the zmet2a methyltransferase nucleic acid sequence.

Definitions

Units, prefixes, and symbols can be denoted in the SI accepted form. Numeric ranges are inclusive of the numbers defining the range. Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation, respectively. The headings provided herein are not limitations of the various aspects or embodiments of the invention which can be had by reference to the specification as a whole. Accordingly, the terms defined immediately below are more fully defined by reference to the specification as a whole.

As used herein, the term "plant" includes reference to whole plants, plant organs (e.g., leaves, stems, roots, etc.), seeds and plant cells and progeny thereof. The class of plants which can be used in the methods of the present invention are generally as broad as the class of higher plants amenable to transformation techniques, including both monocotyledonous and dicotyledonous plants.

As used herein, "heterologous" when used to describe nucleic acids or polypeptides refers to nucleic acids or polypeptides that originate from a foreign species, or, if from the same species, are substantially modified from their original form. For example, a promoter operably linked to a heterologous structural gene is

from a species different from that from which the structural gene was derived, or, if from the same species, one or both are substantially modified from their original form.

A nucleic acid or polypeptide is "exogenous to" an individual plant is one which is introduced into the plant by any means other than by a sexual cross. Examples of means by which this can be accomplished are described below, and include *Agrobacterium*-mediated transformation, biolistic methods, electroporation, and the like. Such a plant containing the exogenous nucleic acid is referred to herein as an R_1 generation transgenic plant. Transgenic plants which arise from sexual cross or by selfing are descendants of such a plant.

As used herein, "zmet2a methyltransferase gene" or "zmet2a methyltransferase nucleic acid" refers to a nucleic acid encoding zmet2a methyltransferase and which hybridizes under stringent conditions and/or has at least 60% sequence identity at the deduced amino acid level to the exemplified sequences provided herein. The zmet2a polypeptide encoded by the zmet2a methyltransferase gene has at least 55% or 60% sequence identity, typically at least 65% sequence identity, preferably at least 70% sequence identity, often at least 75% sequence identity, more preferably at least 80% sequence identity, and most preferably at least 90% sequence identity at the deduced amino acid level relative to the exemplary zmet2a methyltransferase sequences provided herein.

As used herein, "zmet2a methyltransferase nucleic acid" includes reference to a contiguous sequence from a zmet2a methyltransferase gene of at least 2454 nucleotides in length. In some embodiments the nucleic acid is preferably at least 2736 nucleotides in length (see FIG. 1A) and more preferably at least 2796 nucleotides in length (see FIG. 1B).

As used herein, "zmet2b methyltransferase gene" or "zmet2b methyltransferase nucleic acid" refers to a nucleic acid encoding zmet2b methyltransferase and which can be identified using the partial zmet2b methyltransferase nucleic acid shown in FIG. 23. The zmet2b methyltransferase gene

hybridizes under stringent conditions to the partial zmet2b methyltransferase nucleic acid shown in FIG. 23.

As used herein, "a partial zmet2b methyltransferase nucleic acid" includes reference to a contiguous sequence of at least 1181 nucleotides in length and which is from the zmet2b methyltransferase gene.

As used herein, "isolated" includes reference to material which is substantially or essentially free from components which normally accompany or interact with it as found in its naturally occurring environment. The isolated material optionally comprises material not found with the material in its natural environment.

As used herein, "nucleic acid" includes reference to a deoxyribonucleotide or ribonucleotide polymer in either single- or double-stranded form, and unless otherwise limited, encompasses known analogues of natural nucleotides that hybridize to nucleic acids in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence includes the complementary sequence thereof.

As used herein, "operably linked" includes reference to a functional linkage between a promoter and a second sequence, wherein the promoter sequence initiates and mediates transcription of the DNA sequence corresponding to the second sequence. Generally, operably linked means that the nucleic acid sequences being linked are contiguous and, where necessary to joint two protein coding regions, contiguous and in the same reading frame.

In the expression of transgenes, one of ordinary skill in the art will recognize that the inserted nucleic acid sequence need not be identical and may be "substantially identical" to a sequence of the gene from which it was derived. As explained below, these variants are specifically covered by this term.

In the case where the inserted nucleic acid sequence is transcribed and translated to produce a functional zmet2a and/or zmet2b methyltransferase

polypeptide, one of ordinary skill in the art will recognize that because of codon degeneracy, a number of nucleic acid sequences will encode the same polypeptide. These variants are specifically covered by the term "zmet2a methyltransferase nucleic acid sequence" or "zmet2b methyltransferase nucleic acid sequence". In addition, the term specifically includes those full length sequences substantially identical (determined as described below) with a zmet2a and/or zmet 2b methyltransferase gene sequence which encode proteins that retain the function of the zmet2a and/or zmet2b methyltransferase. Thus, in the case of the zmet2a and/or zmet2b methyltransferase genes described herein, the term includes variant nucleic acid sequences which have substantial identity with the sequences disclosed herein and which encode proteins capable of reducing or regulating DNA methylation in a transgenic plant for various purposes as well as silencing target genes in a plant using the nucleic acid sequences described herein.

Two nucleic acids or polypeptides are said to be "identical" if the sequence of nucleotides or amino acid residues, respectively, in the two sequences is the same when aligned for maximum correspondence as described below. The term "complementary to" is used herein to mean that the complementary sequence is identical to all or a specified contiguous portion of a reference nucleic acid sequence. Sequence comparisons between two (or more) nucleic acids or polypeptides are typically performed by comparing sequences of two optimally aligned sequences over a segment or "comparison window" to identify and compare local regions of sequence similarity. Optimal alignment of sequences for comparison may be conducted by the local homology algorithm of Smith and Waterman, *Ad. App. Math.* 2: 482 (1981), by the homology alignment algorithm of Neddleman and Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson and Lipman, *Proc. Natl. Acad. Sci.* (U.S.A.) 85:2444 (1988), by computerized implementation of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (hereinafter "GCG"), 575 Science Dr., Madison, WI), or by inspection.

"Percentage of sequence identity" is determined by comparing two optimally aligned sequences over a comparison window, where the portion of the nucleic acid

sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity.

The term "substantial identity" of nucleic acid sequences means that a nucleic acid comprises a sequence that has at least 55% or 60% sequence identity, generally at least 65%, preferably at least 70%, often at least 75%, more preferably at least 80% and most preferably at least 90%, compared to a reference sequence using the programs described above (preferably BESTFIT) using standard parameters. One of ordinary skill in the art will recognize that these values can be appropriately adjusted to determine corresponding identity of proteins encoded by two nucleotide sequences by taking into account codon degeneracy, amino acid sequences for those purposes normally means sequence identity of at least 55% or 60%, preferably at least 70%, more preferably at least 80%, and most preferably at least 95%. Polypeptides having "sequence similarity" share sequences as noted above except that residue positions which are not identical may differ by conservative amino acid changes. Conservative amino acid substitutions refer to the interchangeability of residues having similar side chains. For example, a group of amino acids having aliphatic side chains is glycine, alanine, valine, leucine, and isoleucine; a group of amino acids having aliphatic-hydroxyl side chains is serine and threonine; a group of amino acids having aromatic side chains is phenylalanine, tyrosine, and tryptophan; a group of amino acids having basic side chains is lysine, arginine, and histidine; and a group of amino acids having sulfur-containing side chains is cysteine and methionine. Preferred conservative amino acids substitution groups are: valine-leucine-isoleucine, phenylalanine-tyrosine, lysine-arginine, alanine-valine, and asparagine-glutamine.

Another indication that nucleic acid sequences are substantially identical is if two molecules hybridize to each other under appropriate conditions. Appropriate

conditions can be high or low stringency and will be different in different circumstances. Generally, stringent conditions are selected to be about 5°C to about 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH 0) at which 50% of the target sequence hybridizes to a perfectly matched probe. Typically, stringent wash conditions are those in which the salt concentration is about 0.22 molar at pH 7 and the temperature is at least about 50°C. However, nucleic acids which do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical. This may occur, e.g., when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code.

Nucleic acids of the present invention can be identified from a cDNA or genomic library prepared according to standard procedures and the nucleic acids disclosed here used as a probe. For example, stringent hybridization conditions will typically include at least one low stringency wash using 0.3 molar salt (e.g., 2X SSC) at 65°C. The washes are preferably followed by one or more subsequent washes using 0.03 molar salt (e.g., 0.2X SSC) at 50°C, usually 60°C, or more usually 65°C. Nucleic acid probes used to isolate the nucleic acids are preferably at least 100 nucleotides in length.

As used herein, a homologue of a particular zmet2a and/or zmet2b methyltransferase gene is a second gene (either in the same species or in a different species) which encodes a protein having an amino acid sequence having at least 50% identity or 75% similarity to (determined as described above) to a polypeptide sequence in the first gene product.

As used herein, "nucleotide binding site" or "nucleotide binding domain" includes reference to a region consisting of kinase-1a, kinase 2, and kinase 3a motifs, which participates in ATP/GTP-binding. Such motifs are described for instance in Yu *et al.*, *Proc. Acad. Sci. USA* 93:11751-11756 (1996); Mindrinos, *et al.*, *Cell* 78:1089-1099 and Shen *et al.*, FEBS, 335:380-385 (1993).

As used herein, "tissue-specific promoter" includes reference to a promoter in which expression of an operably linked gene is limited to a particular tissue or tissues.

As used herein "recombinant" includes reference to a cell, or nucleic acid, or vector, that has been modified by the introduction of a heterologous nucleic acid or the alteration of a native nucleic acid to a form not native to that cell, or that the cell is derived from a cell so modified. For example, recombinant cells express genes that are not found within the native (non-recombinant) form of the cell or express native genes that are otherwise abnormally expressed, under expressed or not expressed at all.

As used herein, a "recombinant expression cassette" is a nucleic acid construct, generated recombinantly or synthetically, with a series of specified nucleic acid elements which permit transcription of a particular nucleic acid in a target cell. The expression vector can be part of a plasmid, virus, or nucleic acid fragment. Typically, the recombinant expression cassette portion of the expression vector includes a nucleic acid to be transcribed, and a promoter.

As used herein, "transgenic plant" includes reference to a plant modified by introduction of a heterologous nucleic acid. Generally, the heterologous nucleic acid is a zmet2a and/or zmet2b methyltransferase structural or regulatory gene or subsequences or combinations thereof.

As used herein, "hybridization complex" includes reference to a duplex nucleic acid sequence formed by selective hybridization of two single-stranded nucleic acids with each other.

As used herein, "amplified" includes reference to an increase in the molarity of a specified sequence. Amplification methods include the polymerase chain reaction (hereinafter "PCR"), the ligase chain reaction (hereinafter "LCR"), the transcription-based amplification system (hereinafter "TAS"), the self-sustained sequence replication system (hereinafter "SSR"). A wide variety of cloning methods,

host cells, and *in vitro* amplification methodologies are well-known to persons of ordinary skill in the art.

As used herein, "nucleic acid sample" includes reference to a specimen suspected of comprising a zmet2a and/or zmet2b methyltransferase gene.

SEQUENCE LISTINGS

The present application contains a number of nucleotide sequences and amino acid sequences. For the nucleotide sequences, the base pairs are represented by the following base codes:

Symbol	Meaning
A	A; adenine
C	C; cytosine
G	G; guanine
T	T; thymine
U	U; uracil
M	A or C
R	A or G
W	A or T/U
S	C or G

Symbol	Meaning
Y	C or T/U
K	G or T/U
V	A or C or G; not T/U
H	A or C or T/U; not G
D	A or G or T/U; not C
B	C or G or T/U; not A
N	(A or C or G or T/U)

The amino acids shown in the application are in the L-form and are represented by the following amino acid-three letter abbreviations:

Abbreviation	Amino acid name
Ala	L-Alanine
Arg	L-Arginine
Asn	L-Asparagine
Asp	L-Aspartic Acid
Asx	L-Aspartic Acid or Asparagine
Cys	L-Cysteine
Glu	L-Glutamic Acid

Gln	L-Glutamine
Glx	L-Glutamine or Glutamic Acid
Gly	L-Glycine
His	L-Histidine
Ile	L-Isoleucine
Leu	L-Leucine
Lys	L-Lysine
Met	L-Methionine
Phe	L-Phenylalanine
Pro	L-Proline
Ser	L-Serine
Thr	L-Threonine
Trp	L-Tryptophan
Tyr	L-Tyrosine
Val	L-Valine
Xaa	L-Unknown or other

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1A shows the nucleic acid sequence of the zmet2a methyltransferase gene containing 2736 basepairs. FIG. 1B shows the nucleic acid sequence of the zmet2a methyltransferase gene containing 2796 basepairs.

FIG. 2A shows the amino acid sequence of the zmet2a methyltransferase containing 912 amino acids and which is encoded by the nucleic acid sequence shown in FIG. 1A. FIG. 2B shows the amino acid sequence of the zmet2a methyltransferase containing 932 amino acids and which is encoded by the nucleic acid sequence shown in FIG. 1B.

FIG. 3 shows the PCR primers used to sequence the zmet2a methyltransferase gene.

FIG. 4 is a summary of the major classes of 5-cytosine methyltransferases from mammals, *Arabidopsis* and maize.

FIG. 5 shows the genomic sequence of zmet2a methyltransferase gene and the retrotransposon SPRITE-1, along with intron-exon divisions, a restriction site map and a primer map.

FIG. 6 lists the World Wide Web sites used to process the sequence data for the zmet2a methyltransferase gene.

FIG. 7 shows a Southern blot of B73 DNA digested with *Hind*III and probed with clone CGET064. The Southern blot shows the presence of multiple copies of zmet2a or zmet2a-like genes in the B73 genome. DNA from B73 was digested with HindIII and probed with clone CGET064 which does not contain a HindIII site. The gene cloned and sequenced is represented by the upper band.

FIG. 8 shows the alignment of the amino acid sequence from zmet2a with the amino acid sequence of *Arbadiopsis* chromosmethylase *CMT1* (AF039367) and the C-terminal methylase domains from the DNA methyltransferases of maize zmet1 (AF063403) and *Arabidopsis MET1* (P34881). Zmet2a shows similarity along the entire length of *CMT1* but significant similarity with zmet1 and Met1 exists only in the conserved motifs. Bold, uppercase, normal uppercase letters, and lower case letters indicate identity, conservation, and differences in amino acid sequences relative to zmet2a respectively. Dashes in the sequences are gaps introduced by CLUSTAL W to optimize the alignments. The location of the six conserved methylase motifs are indicated in the sequence. The chromodomain is located upstream and adjacent to motif IV. The *Mu* insertion into the coding region of motif IX alters zmet2a function resulting in decreased methylation at CpNpG sites. Putative nuclear loalization signal peptides, NLS (N. Raikhel, *Plant Physiol.* 100, 1627 (1992)) are positioned in the N-terminal portion of the protein.

FIG. 9 lists the putative identification of zmet2a amino acids involved in catalysis by comparison with amino acids of M.HhaI with known catalytic functions. The amino acids of M.HhaI with catalytic functions were determined by crystalography by Cheng et al., *Cell*, 74:299-307 (1993). Amino acid of zmet2a are numbered as in Figure 7.

FIG. 10 shows southern analysis of repetitive DNA methylation patterns. Total genomic DNA (5 μg per lane) from an F_4 derived F_5 family segregating for

zmet2a:*Mul* was digested with isoschizomers *Hpa*II and *Msp*I which recoginze the sequence CCGG. Digested DNA ws electophoresed through 0.8% agarose, transferred to nylon membrane, and hybridized with probes for repetitive DNA; the 9kb 26s-5.8s-17s ribosomal repeat (FIG. 10A), 5s ribosomal repeat (FIG. 10B), and a centromeric repeat pSau3a9 (FIG. 10C). Decreased methylation is observed in mutant plants (- -) relative to nonmutant plants (+ +) digested with *Msp*I which is sensitive to methylation at ^{mc}CpCpG sequnces. No changes in methylation patterns at ^{mc}CpG sits are observed in mutant plants as indicated by the lack of digestion with *Hpa*II. Plants heterozygous for zmet2a:*Mul* (+ -) also show decreases at ^{mc}CpCpg sites.

FIG. 11 shows gels from a Southern analysis which demonstrate that plants homozygous for zmet2a::Mu1 have decreased methylation at CpNpG sites. More sites cut with restriction enzymes that are sensitive to methylation at CpNpG sites in zmet2a:*Mul* plants. *Eco*RII is sensitive to methylation at CC*A/TGG sites where * indicates the senitive cytosine (FIG. 11A). *Bgl*II is sensitive to methylation at AGATC*T sites (FIG. 11B). *Pst*I is sensitive to methylation at C*TGCAG sites (FIG. 11C). *Bam*HI is sensitive to methylation at GGATC*C sites (FIG. 11D). *Ava*II is sensitive to methylation at GGA/TC*C sites (FIG. 11E). Changes at CpG sites cannot be sparated from CpCpG in the *Ava*II digests. DNA from the same plants as those in Figure 10 were digested and hybridized with the repetitive probes as described herein.

FIG. 12 shows the cytosine methylation levels in an F4 derived F5 segregating line for zmet2a::Mu1. 5-methylcytosine content of DNA extracted from tissue of immature 5^{th} –7^{th} leaves was determined by reverse phase HPLC using the method of Gehrike et al. Values were obtained from three wildtype plants, seven heterozygous plants and five homozygous plants. Two samples were run for each plant. Percentages of 5mC content [5mC/(5mC + C)] were calculated from concentrations determined from integration of peak and comparison to known standards.

FIG. 13 shows gels from a Southern analysis which demonstrate that plants homozygous for zmet2a::Mu1 having a reduced level of methylation that is stable

over generations. Two F_2 derived F_3 families homozygous for zmet2a:Mul, B5 and B6, were self pollinated to the F_6 generation. Two lineages from B5 and three lineages from B4 were grown at the University of Wisconsin, West Madison Agronomy Farm in 1999. Methylation levels are consistent across generations. Once zmet2a:*Mul* is in a homozygous state, methylation is reduced to a specific level and no further reductions occur. Dilution of methylation is not observed in each successive generation. DNA from leaf tissue was digested with *Msp*I and the Southern blot was hybridized with 9kb ribosomal repetitive probe.

FIG. 14 shows gels from a Southern analysis which demonstrate that methylation levels are restored to nonmutant parental levels in backcross progeny homozygous for wildtype zmet2a. An F1 hybrid of an F4 line homozygous for zmet2a::*Mul* (lanes 1-3) and the inbred line Mo17 (lanes 4-6) was backcrossed to the nonmutant Mo17 parent ot generate plants homozygous wildtype and plants heterozygous for zmet2a:*Mul*. F1 plants (lanes 7-11) have methylation levels intermediate those of the parents. BC1 progeny heterozygous for zmet2a:*Mul* (lanes 12-17) have methylation levels similar to the F1. BC1 plants restored to wild-type zmet2a (lanes 18-21) have remethylation to levels comparable to the nonmutant parent line. Complete or near complete remethylation has occurred within one sexual generation. DNA was extracted from the 4^{th} – 6^{th} immature leaves of greenhouse grown seedlings, digested with *Pst*I which is sensitive to methylation at ^{mc}CTGCAG sequences, and hybridized to the pSau3a9 centromeric repeat.

FIG. 15 shows gels from a Southern analysis which demonstrate the expression of zmet2a in different tissues during development. Southern blots were produced with cDNA's synthesized from mRNA extracted from embryos 24 days after pollination (hereinafter "DAP"), young leaves, immature ear, immature tassel, BMS callus, and 10 day old seedlings. Figure 15A shows the ethidium bromide stained gel. All lanes were loaded with 750 ng of cDNA except for the 10 day seedlings, of which 280 ng was loaded due to the limited amount available. The cDNA's were quantified by spectrophotometry. The marker lane contains 800 ng of lambda DNA digested with *Hind*III. Figure 15B shows the Southern blot hybridized with the zmet2a cDNA probe. Hybridization is observed in tissues that are actively

undergoing cell division. Figure 15C shows the same blot hybridized to a ubiquitin probe to show cDNA loading variation.

FIG. 16 shows the structure of maize retrotransposon SPRITE-1 and sequence of Long Terminal Repeat (hereinafter "LTR") components. FIG. 16A shows that SPRITE-1 consists of long terminal direct repeats, a tRNA primer binding site (hereinafter "PBS"), coding sequence for proteins necessary for replication and transposition, and a polypurine tract (hereinafter "PPT"). FIG. 16B identifies the sequences for the 5' and 3' LTR , PBS and PPT. Each LTR has a 3 base pair inverted repeat which is also shown in the drawing. A putative TATA box is underlined and the putative transcription start site is italicized. The 5 base pair host insertion site duplications are also identified.

FIG. 17 shows the alignments of the conserved protein motifs of the Ty1/copia elements with SPRITE-1. The maize retrotransposon SPRITE-1 is aligned with the retrotransposon hopscotch (U2626) from maizze, retrofit (U72725) from rice, an unpublished *Arabidopsis* retrotransposon (AC006528) and the copia element from Drosophila (M11240).

FIG. 18 shows that the SPRITE-1 copy number and insertion sites differ among maize inbred lines. DNA (7 μg) from inbred maize lines, barley, ice, rye, wheat, and potato was digested with BcoRI which does not cut within the retroelement sequence. The Southern blot was hybridized with a 950 bp SPRITE-1 fragment which includes the 5' untranslated sequence and 5' sequence putatively coding for the *gag* protein but does not include the conserved *gag* motif or the 5' terminal repeat.

FIG. 19 shows the identification of inbred lines containing a SPRITE-1 insertion in zmet2a. PCR was conducted on maize inbred lines from various origins using a primer upstream of the SPRITE-1 insertion site 15F in conjuction with a SPRITE-1 specific primer 18R or a zmet2a primer downstream of the element 8R. The upper panel (15F/18R) show the inbreds that do not have a SPRITE-1 insertion. The lower panel (15F/18R) shows that Mol17 and A682 have a SPRITE-1 insertion

into zmet2a. A682 has an amplification product from both primer sets indicating that it may be hemizygous for SPRITE-1.

FIG. 20 shows expression of retroelement SPRITE-1. Figure 20A shows a Southern blot of cDNAs from roots, immature embryo 24 days after pollination (hereinafter, "DAP"), young leaf, young leaf with inacive zmet2a immature ear, immature tassel, mature pollen, Black Mexican Sweet (hereinafter, "BMS") callus, and 10 day seedling, hubridized with a SPRITE-1 probe. Transcription of SPRITE-1 is evident as indicated by the hybridization to cDNA from embryo, and leaf tissue. Expression is highest in leaf tissue with significantly more expression being observed in leaf tissue from zmet2a:Mul plants that have decreased CpNpG methylation. FIG. 20B shows the same Southern blot hybridized to a ubiquitin probe as a loading control.

FIG. 21 shows that the presence of a SPRITE-1 insertion into a zmet2a intron does not alter transcript splicing. Fragmetns spanning the SPRITE-1 insertion and downstream from the insertion site were amplified by PCR from cDNA's. FIG. 21A shows a scaled representation of zmet2a. Exons are represented by large blocks while the intervening introns are depicted by lines. The insertion of the retroelement is indicated above the zmet2a diagram. The element is inserted in the opposite orientation relative to zmet2a as indicated by the boxed arrows which represent the direct repeats. Positions of the primers used to generate fragments are indicated below the zmet2a diagram. Fragments were amplified from B73 (FIG. 21B) immature ear cDNA which does not contain the retroelement insertion and Mo17 (M) embryo 24 days after pollination cDNA (FIG. 21B) and Mo17 (M) 10 day seedling cDNA (FIG. 21C). No differences were observed on the ethidium bromide stained gel of the PCR products. FIGS. 21B and 21C show hybridization of a near full length B73 cDNA probe to a Southern blot of the PCR fragments.

FIG. 22 shows the methylation status of SPRITE-1. DNA from immature leaves was digested with methylation sensitive restriction enzymes. Southern blots were hybridized with a 970 base pair fragment from the 5' end of the untranslated region of SPRITE-1. There are 5 BstNI/EcoRII sites, 1 MspI/HpaII sites and 1 PstI

site within the sequence context of this probe. Nearly all sites are methylated in this region.

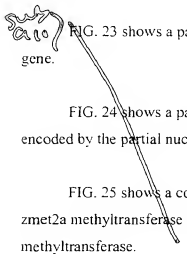

FIG. 23 shows a partial nucleic acid sequence of the zmet2b methyltransferase gene.

FIG. 24 shows a partial amino acid sequence of the zmet2b methyltransferase encoded by the partial nucleic acid sequence shown in FIG. 23.

FIG. 25 shows a comparison of a portion of the amino acid sequence for zmet2a methyltransferase with a portion of the amino acid sequence for zmet2b methyltransferase.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

In one embodiment, the present invention relates to a zmet2a methyltransferase gene. The zmet2a methyltransferase gene of the present invention encodes a class II methyltransferase gene which controls CpNpG methylation. Nucleic acid sequences from the zmet2a methyltransferase gene described herein can be used to reduce or to alter the level of DNA methylation in a plant. In addition, the zmet2a nucleic acid sequence described herein can be used to methylate a targeted gene in a plant *in vivo* to "silence" or "knock-out" said gene.

In another embodiment, the present invention relates a zmet2b methyltransferase gene. The zmet2b methyltransferase gene can be isolated using a partial zmet2b methyltransferase gene described herein. Like the zmet2a methyltransferase gene, the zmet2b methyltransferase gene encodes a class II methyltransferase gene which controls CpNpG methylation. Nucleic acid sequences encoding the zmet2b methyltransferase gene can be used in the same manner as the nucleic acid sequence encoding the zmet2a methyltransferase gene to reduce or to alter the level of DNA methylation in a plant. In addition, the zmet2b nucleic acid sequence can be used to methylate a targeted gene in a plant *in vivo* to "silence" or "knock-out" said gene.

The present invention is applicable to a broad range of types of monocotyledonous and dicotyledonous plants, including, but not limited to, *Zea mays*, *Oryza sativa*, *Secale cereale*, *Triticum aestivum*, *Daucus carota*, *Brassica oleracea*, *Cucumis melo*, *Cucumis sativus*, *Latuca sativa*, *Solanum tubersoum*, *Lycopersicon esculentum*, *Phaseolus vulgaris*, and *Brassica napus*.

The nucleic acids of the present invention can be used in marker-aided selection. Marker-aided selection does not require the complete sequence of the gene or precise knowledge of which sequence confers which specificity. Instead, partial sequences can be used as hybridization probes or as the basis for oligonucleotide primers to amplify by PCR or other methods to follow the segregation of chromosome segments containing the zmet2a and/or zmet2b methyltransferase gene(s) in plants. Because the zmet2a or zmet2b methyltransferase marker is the gene itself, there can be negligible recombination between the marker and the methylated phenotype. Thus, the nucleic acids of the present invention can be used to provide an optimal means to DNA fingerprint class II DNA methyltransferases in other cultivars and wild germplasm. This can be used to indicate if other germplasm accessions and cultivars carry the same zmet2a and/or zmet2b methyltransferase genes.

Preparation of the Nucleic acids of the Present Invention

Generally, the nomenclature and the laboratory procedures involved with recombinant DNA technology described below are those well known and commonly employed by those of ordinary skill in the art. Standard techniques are used for cloning, DNA and RNA isolation, amplification and purification. Generally, enzymatic reactions involving DNA ligase, DNA polymerase, restriction endonucleases and the like are performed according to the manufacturer's specifications. These techniques and various other techniques are generally performed according to Sambrook *et al.*, *Molecular Cloning - A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (1989).

The isolation of zmet2a and/or zmet2b methyltransferase gene(s) can be accomplished via a number of techniques. For instance, oligonucleotide probes based on the sequences disclosed herein can be used to identify the desired gene in a cDNA

or genomic DNA library. To construct genomic libraries, large segments of genomic DNA are generated by random fragmentation, e.g. using restriction endonucleases, and are ligated with vector DNA to form concatemers that can be packaged into the appropriate vector. To prepare a cDNA library, mRNA is isolated from the desired organ of a particular plant, such as shoots from *Zea mays*, and a cDNA library which contains the zmet2a or zmet2b methyltransferase gene transcript is prepared from the mRNA. Alternatively, cDNA may be prepared from mRNA extracted from other tissues in which the zmet2a or zmet2b methyltransferase gene or homologs are expressed.

The cDNA or genomic library can then be screened using a probe based upon the sequence of a cloned zmet2a and/or zmet2b methyltransferase gene or partial sequence from either thereof (such as the partial zmet2b methyltransferase nucleic acid sequence shown in FIG. 23). Probes may be used to hybridize with genomic DNA or cDNA sequences to isolate homologous genes in the same or different plant species.

Those of ordinary skill in the art will appreciate that various degrees of stringency of hybridization can be employed in the assay; and either the hybridization or the wash medium can be stringent. As the conditions for hybridization become more stringent, there is a greater degree of complementarity required between the probe and the target for duplex formation to occur. The degree of stringency can be controlled by temperature, ionic strength, pH and the presence of a partially denaturing solvent such as formamide. For example, the stringency of hybridization is conveniently varied by changing the polarity of the reactant solution through manipulation of the concentration of formamide within the range of 0% to 50%.

Alternatively, the nucleic acids of interest can be amplified from nucleic acid samples using amplification techniques. For instance, polymerase chain reaction (hereinafter "PCR") technology can be used to amplify the sequences of the zmet2a and/or zmet2b methyltransferase and related genes directly from genomic DNA, from cDNA, from genomic libraries or from cDNA libraries. PCR and other *in vitro* amplification methods may also be useful, for example, to clone nucleic acid

sequences that code for proteins to be expressed, to make nucleic acids to use as probes for detecting the presence of the desired mRNA in samples, for nucleic acid sequencing, or for other purposes.

The degree of complementarity (sequence identity) required for detectable binding will vary in accordance with the stringency of the hybridization medium and/or wash medium. The degree of complementarity will optimally be 100 percent; however, it should be understood that minor sequence variations in the probes and primers may be compensated for by reducing the stringency of the hybridization and/or wash medium as described earlier.

Appropriate primers and probes for identifying zmet2a and/or zmet2b methyltransferase nucleic acid sequences from plant tissues are generated from a comparison of the sequences provided herein. For a general overview of PCR see *PCR Protocols: A Guide to Methods and Applications.* (Innis, M, Gelfand, D., Snisky, J. and White, T., eds), *Academic Press*, San Diego (1990), incorporated herein by reference.

Nucleic acids may also be synthesized by well-known techniques as described in the technical literature. See e.g., Curruthers *et al., Cold Spring Harbor Symp. Quant. Biol.* 47:411-418 (1982), and Adams *et al., J. Am. Chem. Soc.* 105:661 (1983). Double stranded DNA fragments may then be obtained either by synthesizing the complementary strand and annealing the strands together under appropriate conditions, or by adding the complementary strand using DNA polymerase with an appropriate primer sequence.

Proteins of the Present Invention

The present invention further provides for isolated zmet2a and/or zmet2b methyltransferases encoded by the zmet2a and/or zmet2b methyltransferase nucleic acids disclosed herein. One of ordinary skill in the art will recognize that nucleic acids encoding a functional zmet2a or zmet2b methyltransferase need not have a sequence identical to the exemplified genes disclosed herein. For example, because of codon degeneracy, a large number of nucleic acid sequences can encode the same

polypeptide. In addition, the polypeptides encoded by the zmet2a and/or zmet2b methyltransferase genes, like other proteins, have different domains which perform different functions. Specifically, zmet2a methyltransferase has ten (10) domains. These ten domains are identified as follows: I, chromodomain β2, chromodomain β3, IV, VI, VIII, IX and X. The ten domains and their sequence ranges (as shown in SEQ ID NO:2) are listed below in Table 1:

TABLE 1

Domain	Amino Acid Sequence Range
I	244-271
Chromodomain β2	366-379
Chromodomain β3	380-388
IV	411-434
VI	456-476
VIII	496-520
IX	723-746
X	751-775

Domains I and X are involved in binding AdoMet, which is source of the methyl group to be transferred during DNA methylation. Domain IV contains a catalytic domain. Domain VI aids in the positioning of domain IV. Domain VIII aids in DNA binding by neutralizing the charge of the phosphodiester backbone. The region between domain VIII and domain IX defines the sequence specificity of the zmet2a methyltransferase enzyme. Thus, the zmet2a methyltransferase gene sequences need not be full length, so long as the desired functional domain of the protein is expressed.

The zmet2a methyltransferase protein is at least 912 amino acid residues in length (see FIG. 2A), preferably, 932 amino acid residues in length (see FIG. 2B). However, those of ordinary skill in the art will appreciate that amino acid deletions, substitutions, or additions to the zmet2a methyltransferase protein will typically yield a enzyme possessing methylating characteristics similar or identical to that of the full length sequence. Thus, full length zmet2a methyltransferase proteins modified by 1,

2, 3, 4, or 5 deletions, substitutions, or additions, generally provide an effective degree of methylation relative to the full-length protein.

A partial amino acid sequence of the zmet2b methyltransferase protein is provided for in FIG. 24 and is 256 amino acids in length.

Modified protein chains can also be readily designed utilizing various recombinant DNA techniques well known to those of ordinary skill in the art. For example, the chains can vary from the naturally occurring sequence at the primary structure level by amino acid substitutions, additions, deletions, and the like. Modification can also include swapping domains from the proteins of the present invention with related domains from other class II methyltransferases.

The present invention also provides antibodies which specifically react with the zmet2a and/or zmet2b methyltransferase(s) of the present invention under immunologically reactive conditions. An antibody immunologically reactive with a particular antigen can be generated *in vivo* or by recombinant methods such as by selection of libraries of recombinant antibodies in phage or similar vectors. The term "immunologically reactive conditions" as used herein, includes reference to conditions which allow an antibody, generated to a particular epitope of an antigen, to bind to that epitope to a detectably greater degree than the antibody binds to substantially all other epitopes, generally at least two times above background binding, preferably at least five times above background. Immunologically reactive conditions are dependent upon the format of the antibody binding reaction and typically are those utilized in immunoassay protocols.

The term "antibody" as used herein, includes reference to an immunoglobulin molecule obtained by *in vitro* or *vivo* generation of the humoral response, and includes both polyclonal and monoclonal antibodies. The term also includes genetically engineered forms such as chimeric antibodies (e.g., humanized murine antibodies), heteroconjugate antibodies (e.g., bispecific antibodies), and recombinant single chain Fv fragments (scFv). The term "antibody" also includes antigen binding forms of antibodies (e.g., Fab^1, $F(ab^1)_2$, Fab, Fv, and, inverted IgG. See, Pierce

Catalog and Handbook, 1994-1995) Pierce Chemical Co., Rockford, IL). An antibody immunologically reactive with a particular antigen can be generated *in vivo* or by recombinant methods such as selection of libraries of recombinant antibodies in phage or similar vectors (*See, e.g.* Huse *et al.*, (1989) *Science* 246:1275-1281; and Ward, *et al.*, (1989) *Nature* 341:544-546; *and Vaughan* et al., (1996) *Nature Biotechnology*, 14:309-314).

Many methods of making antibodies are known to persons of ordinary skill in the art. A number of immunogens are used to produce antibodies specifically reactive to the zmet2a and/or zmet2b methyltransferase(s) of the present invention under immunologically reactive conditions. An isolated recombinant, synthetic, or native zmet2a and/or zmet2b methyltransferase(s) of the present invention is the preferred immunogens (antigen) for the production of monoclonal or polyclonal antibodies.

The zmet2a and/or zmet2b methyltransferase(s) is then injected into an animal capable of producing antibodies. Either monoclonal or polyclonal antibodies can be generated for subsequent use in immunoassays to measure the presence and quantity of the zmet2a and/or zmet2b methyltransferases. Methods of producing monoclonal or polyclonal antibodies are known to those of skill in the art (*See,* Coligan (1991) *Current Protocols in Immunology* Wiley/Greene, NY; and Harlow and Lane (1989) *Antibodies: A Laboratory Manual* Cold Spring Harbor Press, NY); Goding (1986) *Monoclonal Antibodies: Principles and Practice* (2d ed.) Academic Press, New York, NY).

Frequently, the zmet2a and/or zmet2b methyltransferase(s) and antibodies will be labeled by joining, either covalently or non-covalently, a substance which provides for a detectable signal. A wide variety of labels and conjugation techniques are known and are reported extensively in both the scientific and patent literature. Suitable labels include radionucleotides, enzymes, substrates, cofactors, inhibitors, fluorescent moieties, chemiluminescent moieties, magnetic particles, and the like. Patents teaching the use of such labels include U.S. Patent Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149; and 4,366,241.

The antibodies of the present invention can be used to screen plants for the expression of the zmet2a and/or zmet2b methyltransferase(s). The antibodies of the present invention are also used for affinity chromatography in isolating zmet2a and/zmet2b methyltransferase(s).

The present invention further provides zmet2a and/or zmet2b methyltransferase polypeptides that specifically bind, under immunologically reactive conditions, to an antibody generated against a defined immunogen, such as an immunogen consisting of the polypeptides of the present invention. For example, immunogens will generally be at least 912 contiguous amino acids from the zmet2a methyltransferase polypeptide of the present invention. Nucleic acids which encode such cross-reactive zmet2a and/or zmet2b methyltransferase polypeptides are also provided by the present invention. The zmet2a/zmet2b methyltransferase polypeptides can be isolated from any number of plants as discussed earlier. Preferred plants are *Zea mays*, *Oryza* sativa, *Secale cereale*, *Triticum aestivum*, *Daucus carota*, *Brassica oleracea*, *Cucumis melo*, *Cucumis sativus*, *Latuca sativa*, *Solanum tubersoum*, *Lycopersicon esculentum*, *Phaseolus vulgaris*, and *Brassica napus*.

As used herein, the term, "specifically binds" includes reference to the preferential association of a ligand, in whole or part, with a particular target molecule (i.e., "binding partner" or "binding moiety" relative to compositions lacking that target molecule). It is, of course, recognized that a certain degree of non-specific interaction may occur between a ligand and a non-target molecule. Nevertheless, specific binding, may be distinguished as mediated through specific recognition of the target molecule. Typically, specific binding results in a much stronger association between the ligand and the target molecule than between the ligand and non-target molecule. Specific binding by an antibody to a protein under such conditions requires an antibody that is selected for its specificity for a particular protein. The affinity constant of the antibody binding site for its cognate monovalent antigen is at least 10^7, usually at least 10^9, more preferably at least 10^{10}, and most preferably at least 10^{11} liters/mole. A variety of immunoassay formats are appropriate for selecting antibodies specifically reactive with a particular protein. For example, solid-phase

ELISA immunoassays are routinely used to select monoclonal antibodies specifically reactive with a protein (See Harlow and Lane (1988) *Antibodies, A Laboratory Manual*, Cold Spring Harbor Publications, New York, for a description of immunoassay formats and conditions that can be used to determine specific reactivity). The antibody may be polyclonal but preferably is monoclonal. Generally, antibodies cross-reactive to zmet2a and/or zmet2b methyltransferases are removed by immunoabsorbtion.

Immunoassays in the competitive binding format are typically used for cross-reactivity determinations. For example, an immunogenic zmet2a and/or zmet2b methyltransferase polypeptide is immobilized to a solid support. Polypeptides added to the assay compete with the binding of the antisera to the immobilized antigen. The ability of the above polypeptides to compete with the binding of the antisera to the immobilized zmet2a and/zmet2b methyltransferase polypeptides are compared to the immunogenic zmet2a and/or zmet2b methyltransferase polypeptide(s). The percent cross-reactivity for the above proteins is calculated, using standard calculations. Those antisera with less than 10% cross-reactivity with such proteins as zmet2a and/or zmet2b methyltransferase(s) are selected and pooled. The cross-reacting antibodies are then removed from the pooled antisera by immunoabsorbtion with the non-zmet2a and/or non-zmet2b methyltransferase polypeptide(s).

The immunoabsorbed and pooled antisera are then used in a competitive binding immunoassay to compare a second "target" polypeptide to the immunogenic polypeptide. In order to make this comparison, the two polypeptides are each assayed at a wide range of concentrations and the amount of each polypeptide required to inhibit 50% of the binding of the antisera to the immobilized protein is determined using standard techniques. If the amount of the target polypeptide required is less than twice the amount of the immunogenic polypeptide that is required, then the target polypeptide is said to specifically bind to an antibody generated to the immunogenic protein. As a final determination of specificity, the pooled antisera is fully immunoabsorbed with the immunogenic polypeptide until no binding to the polypeptide used in the immunoabsorbtion is detectable. The fully immunoabsorbed antisera is then tested for reactivity with the test polypeptide. If no reactivity is

observed, then the test polypeptide is specifically bound by the antisera elicited by the immunogenic protein.

Production of Recombinant Expression Cassettes

Isolated sequences prepared as described herein can then be used to provide recombinant expression cassettes. One of ordinary skill in the art will recognize that the nucleic acids used in the recombinant expression cassettes described herein encoding a functional zmet2a and/or zmet2b methyltransferase(s) need not have a sequence identical to the exemplified genes disclosed herein. In addition, the polypeptides encoded by the zmet2a and/or zmet2b methyltransferase genes, like other proteins, have different domains which perform different functions. Thus, the zmet2a and/or zmet2b methyltransferase gene sequences need not be full length, so long as the desired functional domain of the protein is expressed.

A DNA sequence coding for the desired zmet2a and/or zmet2b methyltransferase polypeptide(s), for example a cDNA or a genomic sequence encoding a full length protein, can be used to construct a recombinant expression cassette which can be introduced into a desired plant. An expression cassette will typically comprise the zmet2a and/or zmet2b methyltransferase nucleic acid(s) operably linked in either the sense or antisense direction to transcriptional and translational initiation regulatory sequences which will direct the transcription of the sequence from the zmet2a and/or zmet2b methyltransferase gene(s) in the intended tissues for the transformed plant.

For example, a plant promoter fragment may be employed which will direct expression of the zmet2a and/or zmet2b methyltransferase in all tissues of a regenerated plant. Such promoters are referred to herein as "constitutive" promoters and are active under most environmental conditions and states of development or cell differentiation. Examples of constitutive promoters includes the cauliflower mosaic virus (CaMV) 35S transcription initiation region, the 1' or 2' - promoter derived from T-DNA of *Agrobacterium tumefaciens*, and ubiquitin other transcription initiation regions from various plant genes known to those of ordinary skill in the art.

Alternatively, the plant promoter may direct expression of the zmet2a and/or zmet2b methyltransferase gene in a specific tissue or may be otherwise under more precise environmental or developmental control. Such promoters are referred to here as "inducible" promoters. Examples of environmental conditions that may effect transcription by inducible promoters include pathogen attack, anaerobic conditions, or the presence of light.

Examples of promoters under developmental control include promoters that initiate transcription only in certain tissues, such as leaves, roots, fruit, seeds, or flowers. The operation of a promoter may also vary depending on its location in the genome. Thus, an inducible promoter may be fully or partially constitutive in certain locations.

The endogenous promoters from the zmet2a and/or zmet2b methyltransferase genes of the present invention can be used to direct expression of the genes. These promoters can also be used to direct expression of heterologous structural genes. The promoters can be used, for example, in recombinant expression cassettes to drive expression of genes to produce DNA methyltransferase in a particular cell or tissue.

To identify the promoters, the 5' portions of the clones described herein are analyzed for sequences characteristic of promoter sequences. For instance, promoter sequence elements include the TATA box consensus sequence (TATAAT), which is usually 20 to 30 base pairs upstream of the transcription start site. In plants, further upstream from the TATA box, at positions -80 to -100, there is typically a promoter element with a series of adenines surrounding the trinucleotide G (or T) N G. J. Messing et al., in *Genetic Engineering in Plants*, pp. 221-227 (Kosage, Meredith and Hollaender, eds. 1983).

If proper polypeptide expression is desired, a polyadenylation region at the 3'-end of the zmet2a or zmet2b methyltransferase coding region should be included. The polyadenylation region can be derived from the natural gene, from a variety of other plant genes, or from T-DNA.

The vector comprising the sequences from the zmet2a and/or zmet2b methyltransferase gene(s) will typically comprise a marker gene which confers a selectable phenotype on plant cells. For example, the marker may encode biocide resistance, particularly antibiotic resistance, such as resistance to kanamycin, G418, bleomycin, hygromycin, or herbicide resistance, such as resistance to chlorosulforon.

As discussed above, the zmet2a and/or zmet2b methyltransferase gene(s) can be inserted into a recombinant expression cassette in the antisense direction. Expression of the zmet2a and/or zmet2b methyltransferase gene(s) in antisense direction will result in the production of antisense RNA. As is well known, a cell manufactures protein by transcribing the DNA of the gene encoding a protein to produce RNA, which is then processed to messenger RNA (mRNA) (e.g., by the removal of introns) and finally translated by ribosomes into protein. This process may be inhibited in the cell by the presence of antisense RNA. The term antisense RNA means an RNA sequence which is complementary to a sequence of bases in the mRNA in question in the sense that each base (or the majority of bases) in the antisense sequence (read in the 3' to 5' sense) is capable of pairing with the corresponding base (G with C, A with U) in the mRNA sequence read in the 5' to 3' sense. It is believed that this inhibition takes place by formation of a complex between the two complementary strands of RNA, thus preventing the formation of protein. How this works is uncertain: the complex may interfere with further translation, or degrade the mRNA, or have more than one of these effects. This antisense RNA may be produced in the cell by transformation of the cell with an appropriate DNA construct designed to transcribe the non-template strand (as opposed to the template strand) of the relevant gene (or of a DNA sequence showing substantial homology therewith).

The use of antisense RNA to downregulate the expression of specific plant genes is well known. Reduction of gene expression has led to a change in the phenotype of a plant, either at the level of gross visible phenotypic difference (e.g., lack of anthocyanin production in flower petals of petunia leading to colorless instead of colored petals (see van der Krol et al., *Nature*, 333:866-869 (1988)), or at a more subtle biochemical level, for example, a change in the amount of polygalacturonase

and reduction in depolymerization of pectin during tomato fruit ripening (Smith et al., *Nature*, 334:724-726 (1988)). Another more recently described method of inhibiting gene expression in transgenic plants is the use of sense RNA transcribed from an exogenous template to downregulate the expression of specific plant genes (Jorgensen, Keystone Symposium "Improved Crop and Plant Products through Biotechnology", Abstract X1-022 (1994)). Thus, both antisense and sense RNA have been proven to be useful in achieving downregulation of gene expression in plants, which are encompassed by the present invention.

Production of Transgenic Plants

Techniques for transforming a wide variety of higher plant species using the recombinant expression cassettes hereinbefore described are well known and described in the technical and scientific literature. See, for example, Weising *et al.*, *Ann. Rev. Genet.* 22:421-477 (1988).

The hereinbefore described recombinant expression cassettes may be introduced into the genome of a desired plant host by a variety of conventional techniques. For example, the DNA construct may be introduced directly into the genomic DNA of the plant cell using techniques such as electroporation, PEG poration, particle bombardment and microinjection of plant cell protoplasts or embryogenic callus, or the DNA constructs can be introduced directly to plant tissue using ballistic methods, such as DNA particle bombardment. In the alternative, the DNA constructs may be combined with suitable T-DNA flanking regions and introduced into a conventional *Agrobacterium tumefaciens* or *Agrobacterium rhizogenes* host vector. The virulence functions of the *Agrobacterium* host will direct the insertion of the construct and adjacent marker into the plant cell DNA when the cell is infected by the bacteria.

Transformation techniques are known in the art and well described in the scientific and patent literature. The introduction of DNA constructs using polyethylene glycol precipitation is described in Paszkowski *et al.*, *EMBO J.* 3:2712-2722 (1984). Electroporation techniques are described in Fromm *et al.*, *Proc. Natl.*

Acad. Sci. USA 82:5824 (1985). Biolistic transformation techniques are described in Klein *et al.*, *Nature* 327:70-73 (1987).

Agrobacterium tumefaciens-mediated transformation techniques are well described in the scientific literature. See, for example Horsch *et al.*, *Science* 233:496-498 (1984), and Fraley *et al.*, *Proc. Natl. Acad. Sci. USA* 80:4803 (1983). Although *Agrobacterium* is useful primarily in dicots, certain monocots can be transformed by *Agrobacterium*. For instance, *Agrobacterium* transformation of rice is described by Hiei *et al.*, *Plant J.*, 6:271-282 (1994).

Transformed plant cells which are derived by any of the above transformation techniques can be cultured to regenerate a whole plant which possesses the transformed genotype. Such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker which has been introduced together with the zmet2a and/or zmet2b methyltransferase nucleotide sequence(s). Plant regeneration from cultured protoplasts is described in Evans *et al.*, *Protoplasts Isolation and Culture, Handbook of Plant Cell Culture*, pp. 124-176, MacMillian Publishing Company, New York, 1983; and Binding; *Regeneration of Plants, Plant Protoplasts*, pp. 21-73, CRC Press, Boca Raton, 1985. Regeneration can also be obtained from plant callus, explants, organs, or parts thereof. Such regeneration techniques are described generally in Klee *et al.*, *Ann. Ref. of Plant Phys.* 38:467-486 (1987).

The methods of the present invention are particularly useful for incorporating the zmet2a and/or zmet2b methyltransferase nucleic acid(s) into transformed plants in ways and under circumstances which are not found naturally. In particular, the zmet2a and/or zmet2b methyltransferase(s) may be expressed at times or in quantities which are not characteristic of natural plants.

One of ordinary skill in the art will recognize that after the expression cassette is stably incorporated in transgenic plants and confirmed to be operable, it can be introduced into other plants by sexual crossing. Any of a number of standard breeding techniques can be used, depending upon the species to be crossed.

The hereinbefore described expression cassettes can be inserted into a plant in order to reduce or alter the amount of DNA methylation in a plant. Preferably, such an expression cassette contains the zmet2a and/or zmet2b methyltransferase gene(s) inserted into the cassette in the antisense direction as described earlier. A reduction or alteration in the amount of DNA methylation in a plant can be used to stabilize transgene expression in a transgenic plant.

One of the difficulties with the production of transgenic plants is that many transgenes are silenced or are not stable through successive generations. In many cases, transgene silencing is associated with increased DNA methylation. The hereinbefore described expression cassettes of the present invention containing the zmet2a and/or zmet2b methyltransferase gene(s) in the antisense direction can be inserted into a plant either before, concurrently with or after the insertion of another expression cassette containing a transgene which is to be expressed in the plant, such as, but not limited to, a resistance or drought tolerance gene, etc. The antisense RNA produced by the hereinbefore described expression cassette can then form a complex with the endogenous mRNA from the zmet2a and/zmet2b methyltransferase gene(s) within the plant. This complex should reduce or alter the amount of DNA methylation occurring *in vivo* in the plant. This reduction in DNA methylation should prevent the silencing of the desired transgene in the plant.

In a similiar manner, the expression cassettes described herein can be used to modify or alter the yield or biochemical qualities of a plant. As discussed earlier, certain genes in plants and animals are expressed differentially when transmitted thorough a male versus female parent. This phenomenon is known as imprinting. Imprinting is an epigenetic system correlated with DNA methylation. A reduction or alteration of DNA methylation in a plant by transforming a plant with an expression cassette containing the zmet2a and/or zmet2b methyltransferase gene(s) in the antisense direction may affect the yield and biochemical qualities of a plant.

The hereinbefore described expression cassettes can also be used to silence the expression of a particular targeted gene in plants *in vivo*. More specifically, the

expression cassettes of the present invention containing a tissue-specific promoter and the zmet2a and/or zmet2b methyltransferase gene(s) in the sense direction can be inserted into a plant. The tissue-specific promoter will direct expression of the zmet2a and/or zmet2b methyltransferase gene(s) in a area containing the desired targeted gene. Translation of the zmet2a and/or zmet2b methyltransferase gene(s) in the specific area will result in an increase in methylation in the area of the targeted gene. This increase in methylation can silence the targeted gene.

Transgenic plants containing the expression cassettes described herein and which exhibit a reduction in DNA methylation can be identified by using methylation sensitive restriction enzymes or High Performance Liquid Chromatography. Techniques for using methylation sensitive restriction enzymes and High Performance Liquid Chromatography are well known in the art. Transgenic plants containing the expression cassettes described herein and which exhibit an increase in DNA methylation can be identified by using a Northern Blot analysis which is well known in the art.

Additionally, the hereinbefore described expression cassettes can be used in gene therapy for human diseases which are caused by the amplification of trinucleotide repeats.

The following Examples are offered by way of illustration, not limitation.

EXAMPLES

EXAMPLE 1 -Cloning and Sequencing of Zmet2a

a. Cloning and Sequencing

A partial cDNA clone (CGET064) from an immature tassel cDNA library was obtained from Pioneer Hi-Bred International (Des Moines, Iowa). This clone was identified in an expressed tag sequence (hereinafter "EST") database using known DNA methyltransferase sequences for comparison. This original cDNA clone contained sequences from bp 151 to bp 2569 shown in FIG. 1A and 1B. The sequence of this clone, which represents the 3' end of the transcript was used to design forward and reverse primers for 5' and 3' Rapid Amplification of cDNA Ends

(hereinafter "RACE"). RACE was conducted using the Marathon cDNA Amplification Kit (available from Clontech) on cDNA prepared from Mo17 10 day old seedling mRNA. Mo17 is publically available from the National Seed Storage Lab (Fort Collins, Colorado). RACE products were isolated and ends sequenced using Marathon primers and gene specific primers. The remaining sequence was obtained from PCR products by primer walking. The primers used were AP2, 1F, 1R, 2R, 3R, 4F, 5F, 8R, 8F, 9R, 9F, 14F, 17F, and RaceRT (see FIG. 3). Two sequencing passes were made on the Mo17 cDNA ends and four sequencing passes were made on the intervening regions, three from Mo17 cDNA and one from B73. B73 is publically available from the National Seed Storage Lab (Fort Collins, Colorado). A consensus sequence for the coding region was generated and is shown in FIG. 1A and 1B.

Genomic sequence spanning primers 1F and 1R were obtained from Pioneer Hi-Bred International. To obtain the remaining genomic sequence of zmet2a, the CGET064 clone was used to probe a Mo17 genomic library (Stratagene). Lambda clones 4a, 4c, 4d1 and 4d2 were determined to be positive clones containing a sequence identical to CGET064. Lamda clone 4a did not contain the full length gene, therefore, sequence data was obtained from clone 4c. No analysis of clones 4d1 or 4d2 was conducted. Clone 4c was subcloned into pGEM7zf(+) (Promega) using double digests involving *Hind*III, *Xho*I, *Eco*RI, and *Bam*HI. Genomic sequence was obtained from a combination of subclones pHX8 (bp 7311-8878), pHX9 (bp 9173-10135), and pB11(bp 5269-8447) and by primer walking using primers T7, Sp6, M13F, M13R, Seq2FN, Seq2RN, S3F, S3R, 7F, 8eR, 9F, 9R, 11iR, 11iF, 12iR, 12iF, 13iR, 13iF, 14F, 14R, 15R, 15F, 16R, 16F, 17R, 17F, 18R, 18F, and RaceRT (see FIG. 3). Borders of the *Mu* insertion of zmet2a::MU1 were sequenced from PCR products using primer 5F and a *Mu* primer (see FIG. 3). Map locations of the zmet2a primers are shown in FIG. 5.

PCR products were sequenced using Big Dye terminator cycle sequencing on an ABI sequencer (Perkin-Elmer Applied Biosystems) at the University of Wisconsin Biotechnology Center Sequencing Facility (Madison, WI). Sequence data was processed using computional tools available through the World Wide Web (hereinafter, "WWW"), summarized in FIG. 6.

b. Mutant Analysis

A mutant allele called (zmet2a::Mu1) was obtained from Pioneer Hi-Bred International's TUSC system. This mutant allele contains a *Mutator* transposable element insertion and was identified in a *Mutator* population using a *Mu* specific primer and a zmet2a gene specific primer. Since the *Mutator* population is quite variable, heterozygous zmet2a::mu1 F_2 seed was advanced by selfing at the University of Wisconsin West Madison Agronomy Farm (Madison, Wisconsin), the University of Wisconsin Walnut Street greenhouses (Madison, Wisconsin), and at the University of Wisconsin winter nursery in Puerto Rico to produce the F_4 derived F_5 segregating family primarily used in this example.

DNA from 15 plants of the F_4 derived F_5 segregating family was used for HPLC analysis. A subset of these plants was used for Southern analysis. The 5th to 7th immature leaf tips were collected and immediately frozen in dry ice. Tissue was ground in liquid nitrogen and DNA was extracted using a modified CTAB method of Saghai-Maroof et al. (*Proc. Natl. Acad. Sci. USA* 81:8014-8018 (1984)). Tissue was incubated in CTAB (Sigma) extraction buffer for 2 hours at 65 °C, extracted with chloroform/isoamyl alcohol, treated with 0.5 mg RNase A (Sigma) for 30 minutes at 37 °C, extracted again with choroform/isoamyl alcohol, precipitated with isopropanol, washed with 10mM ammonium acetate/76% ethanol, and resuspended in TE.

Plants were genotyped by Southern analysis. DNA (10µg) was digested with *Bam*HI and *Eco*RI which cut on each side of the *Mu* insertion. The digested DNA was electrophoresed through a 0.8% agarose 0.5X TBE gel. DNA was transferred to Immobillon nylon membrane (Millipore) with 5X SSC. Blots were UV cross-linked for 25 seconds and dried at 80 °C for 1.5 hours. Pre-hybridization was carried out in 5X SSC, 50mM Tris pH 8.0, 0.2% SDS, 10 mM EDTA, 2.5X Denhardts solution, and 0.1 mg/ml single stranded sheared herring DNA overnight (8-16 hours) at 65 °C. Hybridization conditions were similar to pre-hybridization except for the addition of 5% dextran sulfate to the hybridization solution. Probes (25-50 ng) (clone CGET064 for genotyping) were radioactively labeled using a random priming reaction

containing 50 μCi of P-32 labeled dCTP. Following overnight hybridization at 65 °C, blots were washed 2X (0.15X SSC, 0.1% SDS) for 30-45 minutes at 65 °C. Hybridized blots were then exposed to Kodak Biomax film.

Southern analysis with methylation sensitive restriction enzymes was conducted in a similar manner except that 5 μg of DNA was digested. Enzymes included in the study were; *Apa*I, *Ava*II, *Bam*HI, *Bgl*II, *Bst*NI, *Cla*I, *Eco*O109, *Eco*RI, *Eco*RII, *Hae*III, *Hin*fI, *Hha*I, *Hpa*II, *Msp*I, *Pst*I, *Pvu*II, *Sac*I, *Sau*3a, *Scr*F1, *Sma*I, *Xho*I. Probes for repetitive sequence regions of the maize genome including a 9 kb clone for the maize 26s-5.8s-17s repeat (reviewed in McMullen et al., *Molecular Analysis of the Nucleolus Organizer Region in Maize. In: Chromosome Engineering in Plants: Genetics, Breeding, and Evoluation*. Gupta PK, Tsuchiya T. (eds). pp. 561-576 (1991)), the 5s ribosomal subunit clone (Mascia et al., *Gene*, 15:7-20 (1981)), and centromere probe pSau3a9 (Jiang et al., *Proc. Natl. Acad. Sci. USA* 93:14210-14213 (1996)) were used to analyze changes in methylation due to zmet2a::Mu1.

HPLC was conducted according to a modified protocol of Gehrke et al., (*J. Chromat.* 301:199-219 (1984)). Duplicate preparations for each of fifteen plants were analyzed. Twenty-five micrograms of DNA was diluted with water to a volume of 50 μl, denatured at 96 °C for 5 minutes and immediately placed on ice. One hundred microliters of 30mM ammonium acetate (pH 5.3), 5 μl of 20mM Zinc Sulfate and 10 μl Nuclease P1 (1mg/ml in 30mM ammonium acetate (pH 5.3) was added and incubated at 37 °C for 2 hours. This reaction cleaves 5' mononucleotides from single stranded DNA. The pH was adjusted with 20 μl of Tris (pH 8.5) and approximately 15 units of Calf Intestinal Alkaline Phosphatase was added and incubated at 37 °C for an additional 2 hours which converts the nucleotides to nucleosides. Samples were frozen at -20°C until HPLC analysis.

HPLC analysis was conducted at the University of Wisconsin Biotechnology Center. A volume of 50 μl was injected into a Brownlee Lab Spheri-5 RP-8 column. Nucleosides were separated with a flow rate of 0.75 ml/min using a gradient program consisting of 30 minutes in buffer A (0.05M Potassium Phosphate pH 4.0, 2.5% methanol), 19 minutes in buffer B (0.05M Potassium Phosphate pH 4.0, 20%

methanol). The column was flushed with 70% methanol for 13 minutes and then re-equilibrated with buffer A for 23 minutes before the injection of the next sample. All samples were analyzed on a Beckman System Gold chromatograph and nucleosides detected at A260nm and A280nm. Nucleoside and nucleotide standards (Sigma) were used to determine nucleoside peak positions and to create a standard curve to determine nucleoside concentration. The ratio of 5-methylcytosine to total cytosine was calculated and statistical analysis conducted using SAS.

To test remethylation as an indication of *de novo* methylase activity, an F_1 hybrid of an F_4 line homozygous for zmet2a::Mu1 and the inbred line Mo17 was backcrossed to the nonmutant Mo17 parent to generate plants homozygous wild-type and plants heterozygous for zmet2a::Mu1. Seedlings of the F_1, the BC_1 progeny, the Mo17 parent and a sib of the F_4 zmet2a::Mu1 parent were grown in the greenhouse and DNA extraction and Southern analysis conducted as previously described. DNA was digested with *Msp*I and *Pst*I and probed with the aforementioned repetitive clones.

c. Expression Analysis

The expression of zmet2a was determined by hybridizing the zmet2a cDNA probe to a Southern blot of cDNA's prepared from different tissues and tissues at different stages of development. Tissues included in this study are embryos 24 days after pollination, 10 day seedlings, immature ear, immature tassel, immature leaf from mutant and nonmutant plants, and roots. Total RNA was extracted using Trizol (Gibco/BRL) according to the manufacture's protocol. The PolyAttract System (Promega) was used to isolate mRNA's from all tissues except 10 day seedlings which was isolated using oligo dT cellulose columns (Pharmacia). cDNA was synthesized from the isolated RNA's using Marathon cDNA Amplification Kit (Clontech).

d. Results

zmet2a shares sequence similarity with other DNA methyltransferases

zmet2a is a member of a small gene family. Three cohybridizing bands are observed on a Southern blot of B73 DNA digested with *Hind*III and probed with

clone CGET064 which does not contain a *Hind*III restriction site (see FIG. 7). zmet2a, which maps to the long arm of chromosome 10, is coded on 20 exons with 19 intervening introns (FIG. 5). The inferred protein using the first predicted translation start site located within a consensus Kozak sequence (Kozak, *J. Cell. Biol.*, 115:887-903 (1991)) is composed of at least 912 amino acids with a predicted mass of 101 Kd (Kilodaltons). A protein of this size with an affinity for CpNpG sequences was isolated in *Pisum sativum* by Pradhan and Adams (*Plant J.*, 471-481 (1995)).

Comparisons with *Arabidopsis* chromomethylase, *CMT1*

Sequence of zmet2a (FIG. 1A and 1B) reveals that it lacks the large N-terminal domain found in the maintenance enzymes but does possess the six highly conserved motifs of the C-terminal catalytic domain. Database searches using BLAST (http://www.ncbi.nlm.nih.gov/gov/BLAST/) show that zmet2a has highest sequence homology to the *Arabidopsis* chromomethylase, *CMT1* (see Henikoff and Comai, *Genetics*, 148:307-318 (1998)) with 44% identity, 57% conservation. The N-terminal region is larger in zmet2a; however, there is an additional downstream predicted start site, also within a consensus Kozak sequence, that codes for an enzyme of 809 amino acids which is more similar in size to the most closely related *CMT1* which is composed of 791 amino acids.

Alignments of zmet2a with *CMT1* and the catalytic domains of *Arabidopsis* *MET1* and maize *zmet1* maintenance enzymes show conservation in the important functional motifs I, IV, VI, VIII, IX and X providing evidence that it is indeed a DNA methyltransferase (FIG. 8). zmet2a and *CMT1* are 87% conserved across the defined six conserved domains, as shown in the underlining in FIG. 8. Zmet2a and CMT1 also have 60% conservation in the variable region sequence between the defined underlined motifs VIII and IX in FIG. 8, which contains a region known as the target recognition domain in the bacterial methyltransferases. The bacterial methylase M.*Hha*I has been crystalized and functions deduced for the conserved amino acids (Cheng et al., *Cell*, 74:299-307 (1993)). The zmet2a amino acids involved in catalysis were predicted by comparison to M.HhaI. The amino acids interacting with SAM and with cytosine are summarized in FIG. 9.

zmet2a mutant plants have reduced methylation at CpNpG sites

A reverse genetics approach was used to ascertain the function of zmet2a. A F_2 family segregating for a *Mutator (Mu)* insertion in the exon encoding motif IX was identified using a PCR primer for *Mu* and a gene-specific primer for zmet2a. This allele is called zmet2a::Mu1. The insertion of *Mu* into exon 19 results in a transcript that would code for a protein truncated at the point of the *Mu* insertion in motif IX due to the introduction of a stop codon. The resulting protein is expected to be dysfunctional since it lacks Motif X which is required for S-Adenosyl methionine (hereinafter "SAM") binding (Cheng et al. *Cell*, 74:299-307 (1993)).

Reduced methylation observed by restriction enzyme analysis

To reduce the genetic background variation associated with the heterogeneous origin of the *Mutator* population, restriction enzyme analysis was conducted on a F_4 derived F_5 family segregating for zmet2a::Mu1. Restriction enzyme isoschizomers *Hpa*II/*Msp*I in addition to other methylation sensitive enzymes were used to determine methylation pattern differences among the three genotypic classes. *Hpa*II and *Msp*I both recognize the sequence CCGG but differ in their sensitivity to methylation. *Hpa*II digestion is inhibited unless both cytosines are unmethylated whereas *Msp*I can digest $C^{me}CGG$ sequences but not $^{me}CCGG$ sites. The methylation status at CpG sites can be accessed by digesting with *Hpa*II and similarly *Msp*I digestion is used to determine the state of methylation at CpCpG sites specifically and may provide a general indication of methylation changes occurring at CpNpG sites.

Results indicate significant reductions in cytosine methylation at ^{me}CCG sites as indicated by a more complete digestion by *Msp*I in plants homozygous for zmet2a::Mu1 (FIG. 10 A-C). Plants heterozygous for zmet2a::Mu1 were intermediate in their digestion pattern. Although the frequency of methylated cytosines is much higher at CpG sequences, no changes in methylation were observed among the genotypic classes when digested with *Hpa*II (FIG. 10 A-C).

Isoschizomers, *Bst*NI and *Eco*RII recognize the sequence CC(A/T)GG. *Bst*NI is not sensitive to cytosine methylation and *Eco*RII is inhibited at $C^{mc}C(A/T)GG$ sites. Nearly all of these sites are methylated in repetitive sequences as a low level of

*Eco*RII digestion is observed only in zmet2a::Mu1 plants (See FIG. 11), whereas digests with *Bst*NI are completely digested to lower molecular weight fragments for all genotypes. These methylated sites may not be subject to zmet2a activity but may instead be methylated by another member of the zmet2a gene family or by *zmet1* or possibly *de novo* methylated after each cell cycle by *zmet3*. Other restriction enzymes were used to clarify the apparent sequence specificity of methylation reduction at CpNpG sites. As with the isoschizomers, no digestion differences are observed with CpG sensitive enzymes *Hha*I [G^{me}CGC] and *Cla*I [AT^{me}CGAT]. More complete digestion is observed in plants homozygous for zmet2a::Mu1 with enzymes sensitive to methylation at CpNpG sites. FIG. 12 shows digestion patterns for enzymes sensitive to methylation at CpNpG sites: *Eco*RII, *Bgl*II, *Pst*I, *Bam*HI, and *Ava*II. In addition to *Eco*RII as previously mentioned, reduced methylation in one or more of the repetitive regions was observed with *Bgl*II [AGAT^{me}CT], *Pst*I [^{me}CTGCAG], *Bam*HI [GGAT^{me}CC], and *Ava*II [GG(A,T)^{me}C^{me}C]. It should be noted that *Ava*II may include some CpG overlapping sites. Subtle differences in digestion patterns of one or more of the repetitive sequences were also observed with *Sau3a*I [GAT^{me}C], *Apa*I [GGG^{me}CC^{me}C], and XhoI [^{me}CT^{me}CGAG]. With these enzymes it is not possible to unambiguously determine whether the source of the difference is CpG or CpNpG methylation. Differences were also observed with *Scr*FI [C^{me}CNGG] which duplicates the targeted sequences and methylation sensitivities of *Eco*RII, *Msp*I and *Hpa*II. Although in many cases the observed reduction in CpNpG or CpN methylation is minimal, any cases of reduced methylation that could be unambiguously attributed to CpG sites have not been observed.

Reduced methylation observed by HPLC

To further assess the extent of methylation reduction caused by the zmet2a::Mu1 allele, HPLC was used to determine the proportion of methylated cytosines in the same F_5 plants used for restriction enzyme analysis. An 11.6% decrease in 5-methylcytosine was observed in plants homozygous for zmet2a::Mu1 relative to siblings homozygous for wild-type zmet2a (FIG. 12). Heterozygotes were intermediate in 5-methylcytosine content. Differences between the genotypic classes are statistically significant at $\alpha < 0.0001$. Since most methylation is found at CpG sites (Gruenbaum et al., *Nature*, 292:860-862 (1981)), a 12% decrease in the total 5-

methylcytosine content likely accounts for a substantial reduction in methylation at CpNpG sites if the reductions are confined to these sequences.

Several generations of inbreeding does not reduce methylation levels beyond that which is observed in the F_2 homozygous mutant (FIG. 13). In addition, it was also observed that plants restored to a normal zmet2a genotype from zmet2a::Mu1 heterozygotes appeared to have near normal levels of methylation.

Methylation is restored after segregation away from zmet2a::Mu1

To test remethylation, a nonmutant line, Mo17, was crossed to a homozygous mutant line, the resulting F_1 was then backcrossed to the nonmutant Mo17 parent line. Restriction enzyme analysis of backcross progeny show all individuals without the *Mu* insertion have remethylated to levels similar to the backcross parent (see FIG. 14). The increased levels of methylation observed in normal BC_1 progeny appear to be higher than that expected from the segregation of normal Mo17 derived chromosome segments and low methylation mutant segments, which would result in a pattern intermediate between the F_1 and the nonmutant parent. These results indicate either that zmet2a has *in vivo de novo* activity and is responsible for establishing CpNpG methylation patterns, or that a separate *de novo* methyltransferase functions only early in development and that zmet2a is responsible for maintaining these patterns. These results on remethylation are in contrast to those of the reduced methlyation patterns of *Arabidopsis* mutants. Backcross progeny, lacking an antisense *MET1* transgene or the *ddm1* mutation, derived from mutant plants outcrossed to normal plants showed very slow remethylation and required several generations to restore methylation to normal levels (Ronemus et al., *Science*, 273:654-657 (1996), Vongs et al., *Science*, 260:1926-1928 (1993), Kakutani et al., *Genetics*, 151:831-838 (1999)). Similar results were observed in selfed progeny from hemizygous antisense *Met1* plants that did not inherit the transgene (Finnegan et al., *Proc. Natl. Acad. Sci. USA* 93:8449-8454 (1996)) however a centromeric region and some single copy sites did remethylate in the first generation (Finnegan et al., *Annu. Rev. Plant Physiol. Plant Mol. Bio.*, 49:223-247 (1998)).

Other DNA methyltransferases that lack the large N-terminal domain have been presumed to be *de novo* enzymes, however, evidence remains insufficient. *In vitro* expression of *Dnmt*3a and *Dnmt*3b (Okano et al., *Nature Genetics*, 19:219-220 (1998)) did not show a specific preference for hemimethylated DNA or nonmethylated DNA and in vivo expression in *Drosophila* (Lyko et al., *Nature Genet.*, 23:363-366 (1999)) further confirm *de novo* activity, whereas *Dnmt*2 (Okano et al., *Nucleic Acids Res.*,26:2536-2540 (1998)) was shown not to effect *de novo* or maintenance methylation in mice. *Masc1*, in *ascobolus*, is purported to have *de novo* activity through its effect on methylation induced premeiotically (MIP) (Malagnac et al., *Cell*, 91:281-290 (1997)). Another Ascobolus methyltransferase *Masc2* was found to be dispensible for maintenance and de novo methylation *in vivo* (Malagnac et al, *Mol. Micro.* 3:331-338 (1999)).

A chromodomain is present in zmet2a

A distinguishing feature of zmet2a, like *CMT1*, is the presence of the chromodomain. Chromodomains have been demonstrated to target proteins to heterochromatic regions and may also be a site of protein-protein interactions (reviewed by Cavalli and Paro, *Curr. Op. Cell Biol.*, 10:354-360 (1998)). The presence of the chromodomain in zmet2a and *CMT1* potentially suggests targeting of the methyltransferase to chromatin complexes or a role of the methyltransferase in chromatin formation and stability. Furthermore, the observation that zmet2a affects CpXpG methylation may also implicate protein targeting through the chromodomain and targeting of methylation patterns. Stable transcriptionally active or silent states may be determined by the formation of chromatin complexes. The mechanisms involved in the formation of silencing complexes remain unknown. However, there is evidence of the involvement of methylation in transcriptionally silenced states which involve methylation binding proteins, transcriptional repressor complexes, and histone deacetylases (Nan et al., *Nature*, 393:386-389 (1998), Wade et al, *Nature Gen.*, 23:62-66 (1999), Ng et al., *Nature. Gen.* 23:58-61 (1999)).

zmet2a is expressed throughout plant development. Expression is higher in the rapidly dividing tissues of seedling, immature ear and embryos (FIG. 15) consistent with the role of methyltransferases in methylating newly synthesized DNA.

Low expression of zmet2a in terminal tissue (leaves) could serve a protective function against invading DNA if this enzyme does have a *de novo* function.

Example 2 – Cloning and Sequencing of the maize retrotransposon SPRITE-1

This example describes the cloning and sequencing of a maize retrotransposon that is inserted into an intron of zmet2a and is referred to herein as "SPRITE-1".

a. Introduction

Within the genomes of most organisms are DNA elements that can be considered parasitic. These elements confer no phenotype of their own and function only for their propagation and insertion elsewhere in the genome. There are two major classes of these elements based on the mechanisms of propagation. One class propagates using DNA-mediated mechanisms where the element does not code for any polymerase and entirely depends on the replication machinery of the host. This class includes the *Ac*, *Spm*, and *Mu* transposable element systems. The other major class is known as retrotransposons, retrotransposable elements or retroelements (reviewed in Grandbastien, *Trends in Genetics* 8:103-108 (1992); Eickbush, *Origin and Evolutionary Relationships of Retroelements. In The Evolutionary Biology of Viruses* (Morse, S.S., ed).) (1994); Wessler et al., *Current Biology*, 5:814-821 (1995); Bennetzen, *Genome*, 37:565-576 (1996)). These elements are not able to excise from one site and insert into another, as the previously mentioned class is capable, but replicate by an RNA-mediated process. The retroelements code for a reverse transcriptase which is a DNA polymerase that uses RNA as a template.

There are several types of retroelements. The main types are retroviruses, long-terminal-repeat (hereinafter "LTR") retroelements, and non-LTR retroelements. Retroviruses are infectious and have not been found in plants, although one plant LTR-retroelement, SIRE-1 from soybean has coding sequences similar to that of a retoviral envelope protein (Laten et al., *Proc. Natl. Acad. Sci.*, 95:6897-6902 (1998)). The non-LTR class is mainly composed of long interspersed nuclear elements (hereinafter "LINEs") and short interspersed nuclear elements (hereinafter "SINEs"). These elements have been found in plants. Less is known about this class than the others. They do differ from LTR-retroelements in that they contain a poly-A tail at

their 3' end. The LTR-retroelement class has been more extensively described in plants than the other classes of retroelements. The LTR-retroelements are usually categorized as one of two groups based on the similarity with the first elements described in yeast and *Drosophila*. One group shares similarity with the Ty3 elements from yeast and the *gypsy* element of *Drosophila* (Marlor et al., *Mol. Cell. Biol.*, 22:829-846 (1986); Clark et al., *J. Biol. Chem.*, 263:1413-23 (1988)). The other group has similarity with the Ty1 elements of yeast and the *copia* element of *Drosophila*. The element identified in this study is of the Ty1/*copia* class (Clare and Farabaugh, *Proc. Natl. Acad. Sci. USA*, 82:2829-2833 (1985); Mount and Rubin, *Mol. Cell. Biol.* 5:1630-1638 (1985)).

The general structure of a LTR-retroelement is depicted in FIG. 16A. These elements are similar in their structure and replication to retroviruses (reviewed in Witcomb and Hughes, *Ann. Rev. Cell Biol.*, 8:275-306 (1992), Eickbush, *Origin and Evolutionary Relationships of Retroelements.* In *The Evolutionary Biology of Viruses* (Morse, S.S., ed.). New York: Raven Press, pp 121-157 (1994), Bennetzen, *Trends in Microbiology*, 9:347-353 (1996)). These elements have direct repeats at the termini as opposed to the DNA based elements that have inverted terminal repeats. Downstream from the 5' LTR is a primer binding site for a host tRNA that primes the first DNA strand synthesis using reverse transcriptase. One or more open reading frames that code for *gag*, a protease, an integrase, a reverse transcriptase, and RNaseH are located downstream from the primer binding site. After the coding region is a polypurine tract followed by the 3' LTR. Ty3/*gypsy* and Ty1/*copia* elements differ in the postion of the integrase coding region. Ty3/gypsy element have the integrase domain at the end of the coding region whereas Ty1/*copia* element have it positioned between the proteinase and reverse transcriptase regions. The *gag* gene encodes proteins for the nucleocapsid and the highly conserved cysteine-histidine nucleic acid binding domain ($CX_2CX_4HX_4C$). The protease processes the polyprotein into its individual components. The integrase functions to insert a newly replicated element into the host DNA. The reverse transcriptase synthesizes the first DNA strand from the transcribed RNA of the element. The RNase degrades the RNA following first strand synthesis. Retroelements rely on the RNA polymerase of the host for

transcription and the host DNA polymerase for second strand DNA synythesis to complete replication.

Using PCR based methods, retroelements were found within nearly every species of the plant kingdom studied (Flavell et al., *Nuc. Acids Res.* 20:3639-3644 (1992); Voytas et al., *Proc. Natl. Acad. Sci. USA* 89:7124-7128 (1992)). Despite the ubiquitous nature of retroelements, there is great heterogeneity among the element within and among species (Flavell et al., *Nuc. Acids Res.* 20:3639-3644 (1992). Wang et al., *Plant Mol. Biol.*, 33:1051-1058 (1997), Pearce et al., *Mol. Gen. Genet.*, 250:305-315 (1996)).

Retroelements are found to be distributed over the entire lengths of chromosomes in *Avena sativa* (Katsiotis et al., *Genome*, 39:410-417 (1996)) but have also been found to be less abundant in heterochromatin, nucleolar organizer regions, centromeres and telomeres (Pearce et al., *Mol. Gen. Genet.*, 250:305-315 (1996); Moore et al., *Genomics*, 10:469-476 (1991); Aledo et al., *Theor. Appl. Genet.*, 90:1094-1100 (1995); Brandeis et al., *Plant Mol. Biol.*, 33:11-21 (1997)). Retroelement-like sequence were found in centromeric regions of grass chromosomes (Miller et al., *Genetics*, 150:1615-1623 (1998)). Many retroelements were discovered by their associations with plant genes (Johns et al., *EMBO J.*, 4:1093-1102 (1985); Grandbastien et al., *Nature*, 337:376-380 (1989); Camirand et al., *Mol. Gen. Genet.*, 224:33-39 (1990)); White et al., *Proc. Natl. Acad. Sci. USA*, 91:11792-11796 (1994)); Hu et al., *Mol. Gen. Genet.*, 248:471-480 (1995); Bi and Laten, *Plant Mol. Biol.*, 30:1315-1319 (1996), Royo et al., *Mol. Gen. Genet.*, 250:180-188 (1996); Kumekawa et al., *Mol. Gen. Genet.*, 260:593-602 (1999)). Many more retroelements or retroelement fragments have been identified using PCR with degenerate primers (Voytas et al., *Proc. Natl. Acad. Sci. USA*, 89:7124-7128 (1992)); Flavell et al., *Nuc. Acids Res.*, 20:3639-3644 (1992); Flavell et al., *Mol. Gen. Genet.*, 231-233 (1992), Pearce et al., *Mol. Gen. Genet.*, 250:305-315 (1996); Katsiotis et al., *Genome*, 39:410-417 (1996); Wang et al., *Plant Mol. Biol.*, 33:1051-1058 (1997)). Others have been identified through studies for other purposes (Bhattacharyya et al., *Plant Mol. Biol.*, 34:255-264 (1997); Vicient and Martinez-Izquierdo, *Gene*, 184:257-261 (1997);

Manninen and Schulman, *Plant Mol. Biol.*, 22:829-846 (1993)) or by genome sequencing projects.

The Ty3/*gypsy* and the Ty1/*copia* elements can be found in large numbers and may contribute up to 50% of the nuclear DNA of the maize genome (SanMiguel et al., *Science*, 274:765-768 (1996)). A 280 Kb region of the maize genome containing the Adh1-F and u22 genes was composed of retroelements, from 10 different families, inserted within each other. The copy number of Ty1/*copia* elements varies considerably. For example, the Ta1 elements of *Arabidopsis* (Voytas et al., *Genetics*, 126:713-721 (1990)) and the Tst1 element of *Solanum tuberosum* (Camirand et al., *Mol. Gen. Genet.*, 224:33-39 (1990)) have one to only a few copies whereas the maize element PREM-2 (Bennetzen, *Trends in Microbiology*, 9:347-353 (1996)) and the BARE-1 element of *Hordeum vulgare* (Manninen and Schulman, *Plant Mol. Biol.*, 22:829-846 (1993)) may be present at 30,000 or more copies.

The differences in copy number infer differences in expression of retroelements. Retroelements are not expressed at high levels as only a few examples of activity have been observed. The Bs1 and Zeon-1 elements of maize (Johns et al., *EMBO J.*, 4:1093-1102 (1985); Hu et al., *Mol. Gen. Genet.*, 248:471-480 (1995)); the Tos elements of rice (Hirochika et al., *Proc. Natl. Acad. Sci. USA* 93:7783-7788 (1996)) the Tnt1 and Tto1 elements of tobacco (Grandbastien et al., *Nature*, 337:376-380 (1989); Hirochika, *EMBO J.*, 12:2521-2528 (1993)) and the Tnp2 element of *Nicotiana plumbaginifolia* have shown evidence of activity. Retroelement expression is higher in plant tissues under stressful conditions. The Tto1, Tto2 of tobacco and Tos17 element of rice were shown to be activated in tissue culture (Hirochika, *EMBO J.*, 12:2521-2528 1993, Hirochika et al., *Proc. Natl. Acad. Sci.*, USA (1996)). The promotors of the BARE-1 element of barley and the Tnt-1 element of tobacco drove expression of reporter genes in protoplasts (Suoniemi et al., *Plant Mol. Biol.*, 31:295-306 (1996); Pouteau et al., *EMBO J.*, 10:1911-1918 (1991)).

Biotic stresses such as viral, fungal and bacterial infection and abiotic stress such as wounding have also been shown to initiate the expression of Tnt1 and Tto1 retroelements (Pouteau et al., *Plant J.*, 5:535-542 (1994); Moreau-Mhiri et al., *Plant*

J., 9:409-419 (1996); Vernhettes et al., *Plant Mol. Biol.*, 35:673-679 (1997); Mhiri et al., *Plant Mol. Biol.*, 33:257-266 (1997); Grandbastien et al, *Genetica*, 100:241-252 (1997); Takeda et al., *Plant Mol. Biol.*, 36:365-376 (1998)). The Bs1 element of maize may have been mobilized prior to insertion in the Adh1 gene by infection with the barley stripe mosaic virus (Johns et al., *EMBO J.*, 1093-1102 (1985)). Only the expression of BARE-1 has been observed in normal unstressed barley leaves (Suoniemi et al., *Plant Mol. Biol.*, 31:295-306 (1997)).

Under normal conditions, retroelements are transcriptionally inactive and are thus transpositionally inactive. Mechanisms within the host must exist to regulate the activity of the retroelements to prevent potentially deleterious mutations that could occur if retroelement transposition was unchecked. Most retroelements are highly methylated (Bennetzen et al., *Genome*, 37:565-576 (1994)) and possibly in heterochromatic regions and may not be accessible to transcriptional machinery. Though silenced in most cases and active in stressful situations, it has been suggested that retroelement transposition may create mutations that may be of selective advantage and provide a means for adaptation (McClintock, *Science*, 226:792-801 (1984)).

b. Cloning and Sequencing of SPRITE-1.

A zmet2a genomic clone was isolated from a lambda library (Stratagene) constructed from Mo17 genomic DNA. The sequence was obtained from subclones or from PCR products by primer walking. Fragments were sequenced using Big Dye terminator cycle sequencing on an ABI sequencer (Perkin-Elmer Applied Biosystems) at the University of Wisconsin Biotechnology Center Sequencing Facility, Madison, Wisconsin.

Expression analysis was conducted on cDNA's prepared using Marathon cDNA Amplification Kit (Clontech) according to the manufacturer's protocols from mRNA isolated from a Mo17 10 day old seedling, Mo17 immature tassel, B73 immature ear, Black Mexican Sweet (BMS) callus, Mo17 embryo 24 days after pollination, W22 pollen, young roots, and immature leaf tissue from zmet2a normal and mutant plants. Total RNA was extracted using Trizol (Gibco/BRL) according to

manufacturer's protocol. Seedling mRNA was isolated using oligo dT cellulose columns (Pharmacia) all other mRNA isolated using the PolyAttract system (Promega).

c. DNA extraction and Southern analysis for genotyping and methylation analysis.

DNA was extracted from immature leaf blades as described in Saghai Maroof et al. (*Proc. Natl. Acad. Sci. USA* 81:8014-8018 (1984)). The copy number of SPRITE-1 was determined by digesting DNA (10μg) with *Eco*RI which does not cut within the element. The digested DNA was electrophoresed through a 0.8% agarose 0.5X TBE gel. Gels were treated with 0.25N HCl for 15 minutes, denatured in 0.2N NaOH and 0.6 M NaCl for 30 minutes, then neutralized in 0.5 M Tris 1.5 M NaCl for 30 minutes. DNA was transferred to Immobilon nylon membrane (Millipore) with 5X SSC. Blots were dried at 80 °C for 1.5 hours. Pre-hybridization was carried out in 5X SSC, 50 mM Tris pH 8.0, 0.2% SDS, 10 mM EDTA, 2.5X Denhardt's solution, and 0.1 mg/ml single stranded sheared herring DNA overnight (8-16 hours) at 65 °C. Hybridization conditions were similar to pre-hybridization except for the addition of 5% dextran sulfate to the hybridization solution. The blot was probed with a PCR fragment (25-50 ng) amplified from the 5' end of the element. Probes were P-32 (50 μCi) labeled using random priming. Following overnight hybridization at 65 °C, blots were washed 2X (0.15X SSC, 0.1% SDS) for 30-45 minutes at 65 °C. Hybridized blots were then exposed to Kodak BioMax film. Southern analysis with methylation sensitive restriction enzymes was conducted on B73 and Mo17 using the same protocols as for genotyping except that 5 μg of DNA was digested. Enzymes included in the study were the differentially methylation sensitive isoschizomers *Hpa*II/*Msp*I and *Eco*RII/*Bst*NI as well as other methylation sensitive enzymes: *Hha*I, and *Pst*I. Blots were hybridized with probes representing different portions of the element.

d. HPLC analysis.

HPLC was conducted according to a modified protocol of Gehrke et al. (*J. Chromato.*, 301:199-219 (1984)). B73 x Mo17 recombinant inbred lines carrying a SPRITE-1 insertion were determined using PCR with the zmet2a primers 15F and 8R,

and the SPRITE-1 primer 18R. Preparations for each of four plants with and without SPRITE-1 were analyzed. Twenty-five micrograms of DNA was diluted with water to a volume of 50 µl, denatured at 96 °C for 5 minutes and immediately placed on ice. One hundred microliters of 30 mM ammonium acetate (pH 5.3), 5 µl of 20 mM Zinc Sulfate and 10 µl Nuclease P1 (1mg/ml in 30 mM ammonium acetate (pH 5.3) was added and incubated at 37 °C for 2 hours. This reaction cleaves 5' mononucleotides from single stranded DNA. The pH was adjusted with 20 µl of Tris (pH 8.5) and approximately 15 units of Calf Intestinal Alkaline Phosphatase was added and incubated at 37 C for an additional 2 hours which converts the nucleotides to nucleosides. Samples were frozen at -20 °C until HPLC analysis.

HPLC analysis was conducted at the University of Wisconsin Biotechnology Center, Madison, Wisconsin. A volume of 40 µl was injected into a Brownlee Lab Spheri-5 RP-8 column. Nucleosides were separated with a flow rate of 0.75 ml/min using a gradient program consisting of 30 minutes in buffer A (0.05M Potassium Phosphate pH 4.0, 2.5% methanol), 19 minutes in buffer B (0.05M Potassium Phosphate pH 4.0, 20% methanol). The column was flushed with 70% methanol for 13 minutes and then re-equilibrated with buffer A for 23 minutes before the injection of the next sample. All samples were analyzed on a Beckman System Gold chromatograph and nucleosides deteeted at A260 nm and A280 nm. Nucleoside and nucleotide standards (Sigma) were used to determine nucleoside peak positions and to create a standard curve to determine nucleoside concentration. The ratio of 5-methylcytosine to total cytosine was calculated and statistical analysis conducted using SAS.

e. Expression analysis.

The expression of SPRITE-1 was determined by hybridizing a SPRITE-1 probe to a Southern blot of cDNA's prepared from different tissues and tissues at different stages of development. Tissues included in this study are embryos 24 days after pollination, 10 day seedlings, immature ear, immature tassel, immature leaf from mutant and nonmutant plants, roots, BMS callus, and mature pollen. Total RNA was extracted using Trizol (Gibeo/BRL) aecording to the manufacture's protocol. The PolyAttract System (Promega) was used to isolate mRNA's from all tissues except 10

day seedlings which was isolated using oligo dT cellulose columns (Pharmacia). cDNA was synthesized from the isolated RNA's using Marathon cDNA Amplification Kit (Clontech).

f. Results

SPRITE-1 is similar to retrotransposons of the Ty1/*copia* group.

In the process of sequencing the maize methyltransferase gene zmet2a, a retroelement inserted within an intron of this gene was discovered and names SPRITE-1. This element is positioned in opposite transcriptional orientation relative to zmet2a. The insertion spans 5220 bp and possesses all the components of a retroelement. Sequence data indicates that SPRITE-1 is a Long-Terminal-Repeat (hereinafter "LTR") retroelement belonging to the Ty1/*copia* class of retroelements. FIG. 16a depicts the general structural components of SPRITE-1. FIG. 16b shows the sequence of the terminal structural components. SPRITE-1 has a perfect 109 bp direct terminal repeats which includes a 3 bp inverted repeat that flanks the internal element sequence. These repeats have the TG...CA pattern found in most plant retroelements and are also shorter than LTR's of most retroelements. LTR's range in size from 115 bp to 4560 bp from information compiled by Bennetzen (*Trends in Microbiology*, 9:347-353 (1996)). A 5 bp host site duplication flanks the repeats externally. Downstream and adjoining the 5' LTR is a primer binding site (PBS) of 16 bp that has sequence complementary to the wheat germ cytoplasmic initiator methionine tRNA (Ghosh et al., *Nuc. Acids. Res.*, 10:3241-3247 (1982)). Upstream and adjoining the 3' LTR is a polypurine tract of 9 bp. Between the putative transcription start site to the predicted translation start site is a 550 bp untranslated region. SPRITE-1 contains a single open reading frame coding 1485 amino acids ending with the stop codon at the 5' end of the polypurine tract.

Database searches for similar coding sequences using BLAST (http://www.ncbi.nlm.nih.gov/gov/BLAST/) show that SPRITE-1 belongs to a different family of retroelements than any other previously described. The most closely related elements based on overall amino acid similarity include an *Arabidopsis* retroelement (AC006528), Retrofit from *Oryza longstaminata* (U72725), and Hopscotch from *Zea mays* (U12626) all having ~35% identity and ~50%

conservation in amino acid sequence with SPRITE-1. It also shares 29% identity and 45% conservation with the *copia* element from *Drosophila*. No elements were found to have nucleotide similarity with the LTR of SPRITE-1 further indicating that this is a member of a unique family of Ty1/*copia* type elements.

SPRITE-1 has the component retrovirus-like amino acid motifs that code for the proteins necessary for transposition. These motifs are the gag-related protein that contains a Cys-His box also known as the CCHC zinc-binding domain, a protease, an integrase, reverse transcriptase and RNase H. These motifs are ordered as they are in Ty1 and *copia*. FIG. 17 shows amino acid alignments of these conserved region from the similar retroelements previously mentioned. These motifs were similarly positioned relative to each other in these retroelements except the CCHC zinc binding domain which was more variant in position relative to the protease motif. This motif was aligned by hand whereas the alignments of the other motifs were constructed by CLUSTAL W and processed using BOXSHADE. Alignments indicate that SPRITE-1 does possess the component protein coding regions necessary for replication and transposition. The coding regions of many retroelements have shown mutations that create frameshifts or introduce stop codons thus preventing translation of functional proteins and preventing transposition. The coding region of SPRITE-1 is intact and therefore has the potential to transpose.

The number of copies of SPRITE-1 is relatively low but variable.

A survey of inbred lines developed from several different populations and other genetic stocks revealed differences in SPRITE-1 copy number. DNA was digested with *Eco*RI and southern blots hybridized with a probe representing the 5' untranslated region of SPRITE-1. This element does not have any *Eco*RI restriction sites. SPRITE-1 is found at a low copy number in most maize lines. Copy number varies from 3 as in B73 and Mo17 to 5 as in B14 and B79 (FIG. 18). The insertion of SPRITE-1 into zmet2a is only found in Mo17 and not in any other maize inbred line except A682, a line derived from Mo17 (FIG. 19). C.I. 187-2, a Mo17 parental line, does not contain SPRITE-1. This indicates that SPRITE-1 has been active recently, i.e. after the origin of the maize populations used for inbred development.

Expression of SPRITE-1

Expression was investigated by hybridizing a southern blot of cDNAs, sythesized from mRNA from different maize tissues, with a SPRITE-1 probe (FIG. 20). Expression of SRITE-1 was highest in leaf tissue. Expression was highest in leaf tissue from plants with a *MUTATOR* insertion in zmet2a and decreased CpNpG methylation. A low level of expression was observed in most tissues, but this may be due to transcription of other genes containing SPRITE-1 in a sense orientation.

SPRITE-1 does not effect zmet2a transcript processing.

During the sequencing of zmet2a cDNA, no fragments or subclones possessed SPRITE-1 sequence indicating that it is efficiently spliced from the transcript. Aberrant splicing has been observed in genes containing retroelements (Pouteau et al., *Mol. Gen. Genet.*, 228:233-239 (1991), Varagona et al., *Plant Cell*, 4:811-820 (1992), Marillonnet and Wessler, *Plant Cell*, 9:967-978 (1997), Kapitonov and Jurka, *J. Mol. Evol.*, 48:248-251 (1999)). Expression of three alleles of the *waxy* gene of maize was low due to retroelement insertions within introns (Varagona et al., *Plant Cell*, 4:811-820 (1992)). Varagona et al. (*Plant Cell*, 4:811-820 (1992)) found that although the element was spliced out of the *waxy* transcript, long-range splice site recognition was disrupted as exons upstream and downstream of the insertion site were found to be excluded in some transcripts. Further analysis of the *wxG* allele showed tissue specific differences in RNA processing with more correctly spliced transcripts in pollen than in the endosperm (Marillonnet and Wessler, *Plant Cell*, 9:967-978 (1997)).

Alternatively spliced transcripts were searched for by PCR amplification of fragments spanning several exons both upstream and downstream of the SPRITE-1 insertion site. Fragments were amplified from Mo17 seedling and immature embryo cDNA and compared to fragments amplified from B73 immature ear cDNA (FIG. 21). Amplification products were separated on an agarose gel and southern blotted. The Southern blot was hybridized to a near full length zmet2a cDNA. No differences were observed between the B73 and Mo17 products indicating that only correctly spliced fragments were detected. The blot was stripped and probed with retroelement sequences. No transcripts were amplified that contained any SPRITE-1 sequence. In

the tissues examined in this example, no aberrant transcripts were detected. Aberrant splicing products may be at such a low concentration that they are not detectable.

SPRITE-1 does not effect zmet2a expression and function.

Since SPRITE-1 is inserted into an intron of zmet2a, the effect of this insertion on zmet2a activity was investigated. HPLC data shows no methylation differences among the recombinant inbred lines with or without a SPRITE-1 insertion in zmet2a. Lines with a SPRITE-1 insertion had 18.21% ± 1.78 5-methylcytosine whereas lines without the insertion had 18.20% ± 0.24. It is probable that most transcripts are processed correctly since no changes in methylation are observed in plants with a SPRITE-1 insertion.

Regions of SPRITE-1 are hypermethylated

Portions of SPRITE-1 were examined to determine the status of cytosine methylation. Using methylation sensitive restriction enzymes, sites within 970 bp of the untranslated region (hereinafter "UTR") immediately downstream from the transcription start site was analyzed. FIG. 22 shows methylation sensitive restriction digestion patterns for Mo17 and B73. The isoschizomers *Hpa*II and *Msp*I recognize CCGG sequences and are differentially sensitive to methylation. SPRITE-1 has a single *MspI/HpaII* site. Using the SPRITE-1 sequence from Mo17, the zmet2a insertion of SPRITE-1 would generate fragments of 5853 bp and 4625 bp. Other SPRITE-1 insertions would generate fragments of variable lengths. Southern blots show only very large fragments >20 Kb for both *Hpa*II and *Msp*I. *Msp*I does show a smaller fragment size than *Hpa*II but is much larger than the expected size for the zmet2a insertion. This indicates that this site is methylated in most SPRITE-1 copies.

Another pair of isoschizomers *Bst*NI and *Eco*RII recognize the sequence CC(A/T)GG. *Bst*NI is not sensitive to methylation and *Eco*RII will not cut when the internal cytosine is methylated. *Bst*NI should generate SPRITE-1-specific fragments of 6, 54, 135, 252, and 784 bp with the UTR probe. All *Eco*RII fragments were greater than 20 Kb indicating complete methylation of these sites. *Hha*I which recognizes GCGC sites should generate SPRITE-1-specific fragments of 2884 and 257 bp and a zmet2a insertion fragment of 2965 bp. No fragments this small were

observed indicating methylation at these sites. The *PstI* site recognized with this probe was also methylated.

EXAMPLE 2 – Cloning and Sequencing of zmet2b

A lambda library (Stratagene) constructed from Mo17 maize genomic DNA library was screened with the zmet2a methyltransferase nucleic sequences shown in FIG. 1. This screening resulted in the recovery of seven (7) independent clones. Four of these clones corresponded exactly to zmet2a nucleic acid sequence. Another type, represented by only one clone, had limited homology in non-significant regions. Two other clones were very similar to the zmet2a methyltransferase nucleic acid sequence but were definitely not identical to the zmet2a methyltransferase nucleic acid sequence. These clones defined a second gene, referred to as "zmet2b". Primer walking resulted in a partial genomic sequence of zmet2b. Primers specific to zmet2b were designed and used to amplify zmet2b cDNA (using Marathon cDNA Amplification Kit from Clontech according to the manufacturer's protocols). The RACE products were isolated and cloned into p-GEMT-Easy (Promega). Sequence of the RACE products generated a partial cDNA sequence for the 3' end of the gene (see FIG. 23). A partial amino acid sequence encoded by this cDNA sequence is shown in FIG. 24. A comparison of a portion of the amino acid sequences for zmet2a and zmet2b is shown in FIG. 25.

All references cited herein are hereby incorporated by reference.

The present invention is illustrated by way of the foregoing description and examples. The foregoing description is intended as a non-limiting illustration, since many variations will become apparent to those skilled in the art in view thereof. It is intended that all such variations within the scope and spirit of the appended claims be embraced thereby.

Changes can be made to the composition, operation and arrangement of the method of the present invention described herein without departing from the concept and scope of the invention as defined in the following claims.